PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 14 March 15, 1928 Number 3

ON THE SEPARATION OF POINTS OF A CONTINUOUS CURVE BY ARCS AND SIMPLE CLOSED CURVES[1]

By W. L. Ayres

Department of Mathematics, University of Pennsylvania

Communicated February 7, 1928

It has been shown by R. L. Moore[2] that in order that a plane continuum M be a continuous curve it is necessary and sufficient that for every two points, A and B of M, there should exist a subset K of M consisting of a finite number of subcontinua of M and separating A and B in M. A subset K of M is said to *separate A and B in M* if M-K is the sum of two non-vacuous mutually separated sets containing A and B, respectively. In this paper it will be shown that if M is a bounded[3] plane continuous curve, the points A and B can be separated by a set of n arcs[4] of M, where n is the number of components[5] of M-A-B that have both A and B as limit points. To do this it will be shown that there exists a simple closed curve which separates A and B in the plane and which has in common with M only a set of n arcs. We shall make considerable use of the properties of the arc-curves which I have developed in another paper.[6] If K is a subset of a continuous curve M the *arc-curve of K with respect to M* (denoted by $M(K)$) is the point set consisting of all points $[P]$ such that P lies on some arc[4] of M whose end-points belong to K. It follows that if a subset of M separates A and B in the arc-curve of $A + B$ with respect to M, then this subset separates A and B in M. All the point sets considered in this paper are assumed to lie in a two-dimensional Euclidean space E_2.

The theorem does not hold true if A and B are merely closed mutually exclusive subsets of M as may be seen by the following simple example.[7] Let M consist of the circle with center (0, 0) and radius 3 together with the intervals from $(-3, 0)$ to $(3, 0)$, from $(-2, 0)$ to $(0, -2)$ and from $(0, -1)$ to $(0, -3)$. Let the set A consist of the point $(0, -3)$. Let the set B consist of the two points $(0, -1)$ and $(0, 3)$. Then, while M-A-B is connected, there is no single arc of M separating A and B in M.

Lemma. *If J and C are simple closed curves, A_1 and A_2 are points common to J and C, A_1XA_2 is an arc of C such that $<A_1XA_2>$[8] has no point in common with J, BY and DZ are arcs such that (1) $<A_1XA_2>$ contains*

B and D, (2) J-A_1-A_2 contains Y and Z, (3) the sets $<BY>$ and $<DZ>$ contain no point of the set $J + A_1XA_2$, (4) there is a region[9] *R_1 containing B such that every point of $<BY>$ which lies in R_1 lies in the interior of C, (5) there is a region R_2 containing D such that every point of $<DZ>$ which lies in R_2 lies in the exterior of C, then Y and Z separate A_1 and A_2 on the simple closed curve J.*

Proof.—Let R denote the interior or exterior of J according as $<A_1XA_2>$ lies in the interior or exterior of J. Since BY and DZ have only the points Y and Z on J, the domain R contains $BY>$ and $DZ>$. Let R' and R'' be the two domains of $R-<A_1XA_2>$ and let R' be the one whose boundary is the arc A_1YA_2 of J plus the arc A_1XA_2 of C. If Y and Z do not separate A_1 and A_2 on J, the set $<A_1YA_2>$ of J contains the point Z and R' contains $<DZ>$. If p is any point of $<A_1XA_2>$, there exists a region R_p containing the point p but containing no point of the set $J + C - <A_1XA_2>$. If $B = D$, there exists a region R_3 which contains B and is a subset of R_B, R_1 and R_2.[10] There exists an arc B_1WB_2 such that (1) B_1 and B_2 are on $<A_1XA_2>$, (2) $<B_1WB_2>$ is common to R' and R_3, (3) of the two arcs into which B_1 and B_2 divide $J' =$ arc A_1XA_2 of $C +$ arc A_1YA_2 of J, that one which contains B lies in R_3.[11] Then the arc B_1BB_2 of J' is a subset of $<A_1XA_2>$, and B_1 and B_2 separate B and Y and separate D and Z on the simple closed curve J'. Further $<B_1WB_2>$ contains a point P_1 of $<BY>$ and a point P_2 of $<DZ>$.[12] Since $<B_1WB_2>$ is a subset of R' and R_B, $<B_1WB_2>$ contains no point of C. As $<B_1WB_2>$ is contained in R_1 and R_2 the point P_1 belongs to the interior of C and P_2 belongs to the exterior of C. Thus $<B_1WB_2>$ is a connected set containing a point of the interior of C and a point of the exterior of C but containing no point of C. Clearly this is impossible. If $B \neq D$, for any point p of the subarc BD of A_1XA_2 let R'_p be defined as follows: for $p = B$, let R'_p be a region containing B and common to the regions R_1 and R_B;[10] for $p = D$, let R'_p be a region containing D and common to the regions R_2 and R_D;[10] for $B \neq p \neq D$, let R'_p be R_p. For each point p of BD there exists an arc β_p with end-points U_p and V_p on A_1XA_2 in the order $A_1U_pV_pA_2$ such that (1) $<\beta_p>$ is common to R' and R'_p, (2) of the two arcs into which U_p and V_p divide J' that one which contains p lies in R'_p.[11] By the Borel theorem there exists a chain $<U_1V_1>$, $<U_2V_2>$, $<U_3V_3>, \ldots, <U_mV_m>$ of the segments[8] $[<U_ppV_p>]$, such that (1) $<U_1V_1>$ contains B and $<U_mV_m>$ contains D, (2) $<U_iV_i>$ and $<U_jV_j>$ have no points in common unless $i = j \pm 1$. Let k_1 be the largest integer $i (2 \leqq k_1 \leqq m)$, such that $<B_1>$ and $<B_i>$ have a point in common. On the arc β_1 in the order from U_1 to V_1 let T_1 be the first point of $<\beta_{k_1}>$. Let k_2 be the largest integer i such that $<\beta_{k_1}>$ and $<\beta_i>$ have a common point. Let T_2 be the first point of $<\beta_{k_2}>$ on either the subarc $T_1U_{k_1}$ or $T_1V_{k_1}$ of β_{k_1} in the order T_1 to U_{k_1} or T_1 to V_{k_1}. Continue this process.

After a finite number of steps we reach a point T_r which belongs to $<\beta_m>$. Then Λ = subarc U_1T_1 of β_1 + subarc T_1T_2 of β_{k_1} + . . . + subarc $T_{r-1}T_r$ of $\beta_{k_{r-1}}$ + subarc T_rV_m of β_m is an arc having much the same properties as the arc B_1WB_2. We may obtain a contradiction using the arc Λ as we did with the arc B_1WB_2 in the case where $B = D$.

THEOREM 1. *If M is a bounded continuous curve, A and B are distinct points of M and there is just one component of M-A-B that has both A and B as limit points, then M contains an arc which separates A and B in M.*

Proof.—The set $M(A + B)$ is a continuous curve[13] and neither A nor B is a cut-point of $M(A + B)$.[14] Suppose $M(A + B)$ contains a cut-point P. Then $M(A + B) - P$ consists of just two components, one of which contains A and the other contains B.[15] Hence P separates A and B in $M(A + B)$, and thus in M, and is the required arc of the theorem.

Now let us consider the case in which $M(A + B)$ contains no cut-point. In this case the boundary of every complementary domain of $M(A + B)$ is a simple closed curve.[16] Suppose there is a complementary domain D of $M(A + B)$ whose boundary J contains both A and B. Let X and Y be two points of J separating A and B on J. Since there is just one component of M-A-B that has both A and B as limit points, there exists an arc α of M-A-B with end-points X and Y.[17] In the order from X to Y let X' be the last point of the arc AXB of J on α, and let Y' be the first point of the arc AYB of J on the subarc $X'Y$ of α. There is a complementary domain R of M which is a subset of D and such that every boundary point of D is also a boundary point of R.[18] There exists an arc β with end-points X' and Y' and lying in R except for these two points.[19] It is easy to show that β plus the subarc $X'Y'$ of α is a simple closed curve separating A and B in the plane. In this case the subarc $X'Y'$ of α separates A and B in M. For the remainder of the argument we will suppose that there is no complementary domain of $M(A + B)$ whose boundary contains both A and B.

There exists a circle C separating A and B in E_2 and containing at least one point not belonging to M. Let $S_1, S_2, S_3, \ldots$, be the set of all components of the common part of C and E_2-M. There is at least one segment in this set and the set is either finite or countable. Each of the sets S_i is a segment belonging to some complementary domain D_i of $M(A + B)$, and let X_i and Y_i be the end-points of S_i. The domains D_i are not necessarily different for different values of i. Let J_i be the boundary of D_i. At least one of the two arcs of J_i from X_i to Y_i contains neither A nor B since the boundary of no complementary domain of $M(A + B)$ contains both A and B. Let α_i be an arc of J_i with end-points X_i and Y_i and containing neither A nor B. Let

$$G = (C - \sum_{i=1}^{\infty} S_i) + \sum_{i=1}^{\infty} \alpha_i$$

Since only a finite number of the curves J_i are of diameter greater than a preassigned positive number,[19] it follows that G is a continuous curve and it is evident from the definition of G that M contains G. If A and B are not separated in E_2 by G, there exists an arc γ with end-points A and B and containing no point of G. Both cases being alike, we will assume that A is in the interior of C. Let order be defined on γ as being from A to B. Let U_1 be the first point of C on γ. The point U_1 belongs to some segment S_{k_1}. Let V_1 be the last point of S_{k_1} on γ. If the subarc V_1B of γ contains any point of C except V_1, let U_2 be the first such point. There is a segment S_{k_2} ($k_2 \neq k_1$) containing U_2. Let V_2 be the last point of S_{k_2} on γ. After a finite number of steps we reach a point V_m such that the subarc AV_m of γ contains every point of C that lies on γ. Let $V_0 = A$ and $U_{m+1} = B$. Let

$$\eta = \sum_{i=0}^{m} \text{subarc } V_iU_{i+1} \text{ of } \gamma + \sum_{i=1}^{m} \text{arc } U_iV_i \text{ of segment } S_{k_i}.$$

The set η is an arc with end-points A and B and having no point in common with G. Now let order be defined on η as being from A to B. It is evident that there is some integer j ($1 \leqq j \leqq m$) such that the segment $<V_{j-1}U_j>$ of η lies in the interior of C and the segment $<V_jU_{j+1}>$ of η lies in the exterior of C. On the subarc BV_j of η let W_2 be the first point of J_{k_j}, and on the subarc AU_j of η let W_1 be the last point of J_{k_j}. The subarcs U_jW_1 and V_jW_2 of η satisfy the conditions of the Lemma. Hence W_1 and W_2 separate X_{k_j} and Y_{k_j} on the curve J_{k_j}. By definition one of the arcs of J_{k_j} from X_{k_j} to Y_{k_j} belongs to G. This arc α_{k_j} contains one of the points W_1 or W_2. Then η *does* contain a point of G. Therefore, G separates A and B in the plane E_2. Let D be the complementary domain of G containing A. Let L be the outer boundary of D relative to the point B.[20] Then L is a simple closed curve which is a subset of G, and thus of M, and separates A and B in the plane E_2.

Since only a finite number of the curves J_i contain a point whose distance from C is greater than a given positive number,[19] the set L contains an arc A_1ZA_2 which lies on the boundary of some complementary domain D_s of $M(A + B)$. The domain D_s contains a complementary domain D_s' of M whose boundary contains the arc A_1ZA_2.[18] There is an arc A_1WA_2 whose end-points are A_1 and A_2 and which lies in D_s' except for A_1 and A_2.[19] The two cases being alike, we will assume that $<A_1WA_2>$ lies in the complementary domain of L that contains B. Let L_1 be the boundary of the complementary domain of $L + A_1WA_2$ that contains B. The set L_1 is a simple closed curve separating A and B in the plane and having one of the two arcs of L from A_1 to A_2 in common with M. This is the desired arc of M separating A and B in M.

THEOREM 2. *If A and B are distinct points of a bounded continuous*

curve M, there exists a simple closed curve separating A and B in the plane and having in common with M a set of n arcs, where n is the number of components of M-A-B that have both A and B as limit points.

Proof.—I. If $n = 1$, then by theorem 1 there is either a point or an arc that separates A and B in M. If A and B are separated in M by a point P, there exists a simple closed curve separating A and B in E_2 and having only the point P in common with M.[21] In the other case we showed in the proof of theorem 1 the existence of a simple closed curve separating A and B in E_2 and having just one arc in common with M.

II. If $n > 1$, let $H_1, H_2, \ldots, H_n$ denote the set of components of M-A-B that have both A and B as limit points. By theorem 1 each component H_i contains an arc α_i that separates A and B in $H_i + A + B$. There exists a set of n complementary domains of M, $D_1, D_2, \ldots, D_n$ such that (1) both A and B are boundary points of each D_i, (2) each domain D_i has points of just two components of the set $[C_j]$ on its boundary, (3) each component C_j contains boundary points of just two domains of the set $[D_i]$. We may suppose that for each i, C_i and C_{i+1} contain boundary points of D_i.[22] It is easy to see that α_i contains a boundary point q_i of D_i and α_{i+1} contains a boundary point p_{i+1} of D_i. Let β_i be an arc whose end-points are q_i and p_{i+1} and such that D_i contains $<\beta_i>$. Then

$$\sum_{i=1}^{n} (\beta_i + \text{subarc } p_iq_i \text{ of } \alpha_i)$$

is a simple closed curve separating A and B in E_2 and having exactly n arcs in common with M.

[1] Presented to the American Mathematical Society, February 26, 1927. As presented to the Society, this paper contained several results which are not included in it in its present form. I found that one of these results had been previously stated by C. M. Cleveland. See these PROCEEDINGS, **13**, 1927 (275–276). The other results will be published as a part of another paper, "Concerning the Arc-Curves and Basic Sets of a Continuous Curve," which will appear in the *Trans. Amer. Math. Soc.*

[2] *Fund. Math.*, **7**, 1925 (302–307).

[3] This condition can be replaced by the condition that M is not the entire plane and the boundary of every complementary domain of M is bounded.

[4] In this paper a single point is considered as a special case of an arc.

[5] A connected subset H of a point set K is said to be a *component* of K if there is no connected subset of K containing H as a proper subset.

[6] See my paper referred to in footnote 1. Hereafter I shall refer to this paper as "Arc-Curves."

[7] However, it is likely that the theorem will hold under certain restrictions if we assume in addition that A and B are connected.

[8] If XYZ denotes an arc with end-points X and Z, the symbols $<XYZ, XYZ>$ and $<XYZ>$ denote XYZ-X, XYZ-Z and XYZ-X-Z, respectively. By the *segment* XYZ is meant XYZ-X-Z.

[9] We may interpret *region* as the interior of a simple closed curve. See however,

Moore, R. L., *Trans. Amer. Math. Soc.*, **17**, 1916 (131–164). Hereafter we will refer to this paper as "Foundations."

[10] "Foundations," theorem 6.

[11] *Ibid.*, theorem 28.

[12] *Ibid.*, theorem 29.

[13] If K is a closed subset of M, $M(K)$ is a continuous curve. See "Arc-Curves," theorem 7.

[14] "Arc-Curves," theorem 8.

[15] If P is a cut-point of $M(K)$, every component of $M(K)$-P contains at least one point of K. See "Arc-Curves," theorem 9, part (1).

[16] Whyburn, G. T., these PROCEEDINGS, **13**, 1927 (31–38), theorem 10.

[17] Moore, R. L., *Math. Zeit.*, **15**, 1922 (255).

[18] "Arc-Curves," theorem 10, part (5).

[19] Schoenflies, A., *Die Entwicklung der Lehre von den Punktmannigfaltigkeiten*, Zweite Teil, Leipzig, **1908** (237).

[20] Moore, R. L., these PROCEEDINGS, **11**, 1925 (469–476), footnote 5.

[21] See an abstract of a paper by R. G. Lubben, "The Separation of Mutually Separated Subsets of a Continuum by Curves," *Bull. Amer. Math. Soc.*, **32**, 1926 (114).

[22] Subscripts are reduced modulo n.

ON SOME PROPERTIES OF ONE-VALUED TRANSFORMATIONS OF MANIFOLDS

BY H. HOPF

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated February 11, 1928

1. *Some Constants of a Transformation.*—Given any set of ν-cycles in an n-complex μ^n, its maximum number of independent cycles with respect to homologies shall be called the "rank" of the set; it is not greater than the νth Betti number π_ν of μ^n. Consider another n-complex M^n transformed into μ^n by a one-valued continuous transformation f and the set of all $f(c_i^\nu)$, where the c_i^ν are the ν-cycles of M^n; then the rank r_ν of this set is a constant of f ($\nu = 0, 1, \ldots, n$).

If $c_1^\nu, c_2^\nu, \ldots, c_{p_\nu}^\nu$, where p_ν is the νth Betti number of M^n, form a fundamental set in M^n and $\gamma_1^\nu, \gamma_2^\nu, \ldots, \gamma_{\pi_\nu}^\nu$ a fundamental set in μ^n, then the transformations of the ν-cycles define a system of homologies[1]

$$f(c_j^\nu) \approx \sum_{k=1}^{\pi_\nu} \alpha_{jk}^\nu \gamma_k^\nu \quad (j = 1, 2, \ldots, p_\nu) \tag{1}$$

and r_ν is the rank of the matrix

$$A_\nu = \| \alpha_{jk}^\nu \| .$$

When the c_j^ν and γ_k^ν are replaced by other fundamental sets, A_ν is trans-

formed by square matrices whose determinants are equal to ± 1. Therefore, also the invariant factors of the matrices A_ν are constants of f.

If $p_\nu = \pi_\nu$, for instance in the important case where μ^n is identical with M^n, the A_ν are square matrices and their determinants

$$a_\nu = |A_\nu| = |\alpha^\nu_{jk}| \quad (j, k = 1, 2, \ldots, p_\nu)$$

(whose absolute values are the products of the invariant factors), are also constants of f, to be considered.

In this paper we shall deal with geometrical properties of f which are expressible in terms of the constants r_ν and a_ν. We shall assume that M^n and μ^n are *closed connected manifolds*. Therefore,

$$p_0 = p_n = \pi_0 = \pi_n = 1, \quad A_0 = \| 1 \|, \quad A_n = \| a_n \|,$$

where the constant $a_n = a$ is the Brouwer *degree*[2] of f.

2. *A Formula of Contragredience.*—Let

$$L_\nu = \| (c^\nu_i . c^{n-\nu}_j) \|, \quad \Lambda_\nu = \| (\gamma^\nu_i . \gamma^{n-\nu}_j) \|$$

be the intersection matrices of the fundamental sets c^ν_i, $c^{n-\nu}_j$ in M^n and γ^ν_i, $\gamma^{n-\nu}_j$ in μ^n respectively. Their determinants are ± 1,[3] hence the inverses L_ν^{-1}, Λ_ν^{-1} are defined. We shall denote by E_m the matrix unity of order m, and as usual by B' the transverse of a matrix B. We shall prove below (Nos. 6, 7) the following *relation of contragredience*

$$(L^{-1}_{n-\nu} A_{n-\nu} \Lambda_{n-\nu})' A_\nu = a E_{\pi_\nu}, \tag{2}$$

from which all our results will follow at once.

3. *Properties of the Constants r_ν.*—Let first be $a \neq 0$. Then aE_{π_ν} is of rank π_ν; therefore, the rank of no matrix on the left hand of (2) can be $< \pi_\nu$ so that

THEOREM I. *If the degree of $f \neq 0$ then $r_\nu = \pi_\nu$ $(\nu = 0, 1, \ldots, n)$.*

A_ν has p_ν rows; hence

THEOREM Ia. *If, for a certain ν, $p_\nu < \pi_\nu$, then the degree of f must be zero.*

A special case of Ia is

THEOREM Ib. *It is not possible to transform the n-sphere into a manifold of which at least one Betti number $\pi_\nu (\nu = 1, 2, \ldots, n-1)$ is > 0, with a degree different from zero.*

Let us now consider the case $a = 0$. Since the determinants of $L^{-1}_{n-\nu}$ and $\Lambda^{-1}_{n-\nu}$ are $\neq 0$, the matrix $(L^{-1}_{n-\nu} A_{n-\nu} \Lambda_{n-\nu})' = B = \| b_{ij} \|$ has the rank $r_{n-\nu}$. Hence the system

$$\sum_{j=1}^{p_\nu} b_{ij} x_j = 0 \quad (i = 1, 2, \ldots, \pi_\nu)$$

has $\pi_\nu - r_{n-\nu}$ linearly independent solutions $x_{1m}, x_{2m}, \ldots, x_{p_\nu m} (m = 1, 2, \ldots, \pi_\nu - r_{n-\nu})$. Since $a = 0$, according to (2) the elements α^ν_{jk} of A_ν satisfy the equations

$$\sum_{j=1}^{p_\nu} b_{ij}\alpha^\nu_{jk} = 0 \quad (k = 1, 2, \ldots, \pi_\nu).$$

Hence, among the columns $\alpha^\nu_{1k}, \alpha^\nu_{2k}, \ldots, \alpha^\nu_{p_\nu k}$ of A_ν there are at most $\pi_\nu - r_{n-\nu}$ linearly independent; therefore,

THEOREM II. *If f is of degree 0 then $r_\nu + r_{n-\nu} \leqq \pi_\nu$ $(\nu = 0, 1, \ldots, n)$.*

From now on we shall always assume that M^n *and* μ^n *are identical*, although our results will also hold when merely $p_\nu = \pi_\nu$, even if $M^n \neq \mu^n$.

The sets of all ν-cycles of M^n and of all $(n - \nu)$-cycles have the common rank p_ν, so that the sum of their ranks is $2p_\nu = 2\pi_\nu$. Therefore, theorems I and II may now be interpreted in the following manner: *If the transformation f of M^n into itself is of degree $\neq 0$ then there exist no c^ν nor $c^{n-\nu}$ not ≈ 0 with an f-transform ≈ 0. On the other hand, when f is of degree zero then for each ν there exist ν or $(n - \nu)$-cycles not ≈ 0, but whose f-transforms are ≈ 0. The sum of the ranks of these "degenerating" ν- and $(n - \nu)$-cycles is at least half the sum of the ranks of all ν- and $(n - \nu)$-cycles.*

4. *A Relation between the Constants a_ν; an Application.*—The determinants of the matrices L and Λ are always ± 1; under the assumption $M^n = \mu^n$ we have even $|L^{-1}_{n-\nu}| = |\Lambda_{n-\nu}| = \pm 1$, because $L_{n-\nu} = \Lambda_{n-\nu}$. Therefore, by computing the determinants in (2) we find:

THEOREM III. *The constants a_ν and the degree a of a transformation of M^n into itself are related by the equations*

$$a_\nu \cdot a_{n-\nu} = a^{p_\nu} \quad (\nu = 0, 1, \ldots, n).$$

We give an example of an application of this theorem: let M^n be the complex projective plane P. It is a 4-dimensional closed orientable manifold with the Betti numbers

$$p_0 = p_4 = 1, \quad p_1 = p_3 = 0, \quad p_2 = 1.$$

Therefore, III gives $a_2^2 = a$, so that we have

THEOREM IV*a*. *The degree of any transformation of the complex projective plane into itself is a perfect square.*

An example of a transformation with the degree b^2 with arbitrary b is given by $z_i' = z_i^b$ $(i = 1, 2, 3)$, where $z_1 : z_2 : z_3$ represents a point of P and $z_1' : z_2' : z_3'$ its image.

Furthermore, according to the formula of Lefschetz[4] the algebraic number of fixed points under a transformation of P is $1 + a_2 + a_4$; but now we see that this number is equal to $1 + a_2 + a_2^2 \neq 0$ whatever the integer a_2. Therefore, in analogy with a well-known property of the real projective plane,

THEOREM IV*b*. *Every transformation of the complex projective plane has at least one fixed point.*

5. *The Behavior of the Kronecker-Indices under a Transformation.*—In the case $M^n = \mu^n$ (2) may be written

$$(L_{n-\nu}^{-1}\, A_{n-\nu}\, L_{n-\nu})'\, A_\nu = aE_{p_\nu} \tag{2'}$$

or, since $L_{n-\nu} = (-1)^{\nu(n+1)}L_\nu'$ and $(BC)' = C'B'$,

$$L_\nu A_{n-\nu}' L_\nu^{-1} A_\nu = aE_{p_\nu}. \tag{2''}$$

When $a \neq 0$ then also $A_\nu \neq 0$ (Th. I). Hence A_ν^{-1} exists and from (2″) follows $L_\nu A_{n-\nu}' L_\nu^{-1} = aE_{p_\nu} A_\nu^{-1}$. Furthermore, since E_{p_ν} and A_ν are commutative, $L_\nu A_{n-\nu}' L_\nu^{-1} = A_\nu^{-1}.aE_{p_\nu}$,

$$A_\nu L_\nu A_{n-\nu}' = aL_\nu. \tag{3}$$

Consider now the transformations of fundamental sets c_i^ν, $c_j^{n-\nu}$ given by (1):

$$\begin{aligned} f(c_i^\nu) &= \bar{c}_i^\nu \approx \sum_{k=1}^{p_\nu} \alpha_{ik}^\nu c_k^\nu \\ f(c_j^{n-\nu}) &= \bar{c}_j^{n-\nu} \approx \sum_{l=1}^{p_\nu} \alpha_{jl}^{n-\nu} c_l^{n-\nu} \end{aligned} \tag{1'}$$

and the intersection matrix of the transformed cycles $\bar{L}_\nu = \| (\bar{c}_i^\nu . \bar{c}_j^{n-\nu}) \|$. We have from (1′)

$$(\bar{c}_i^\nu . \bar{c}_j^{n-\nu}) = \sum_{k,l} \alpha_{ik}^\nu (c_k^\nu . c_l^{n-\nu}) \alpha_{jl}^{n-\nu},$$

hence,

$$\bar{L}_\nu = A_\nu L_\nu A_{n-\nu}'$$

and in view of (3) the equation

$$\bar{L}_\nu = aL_\nu, \tag{5}$$

from which follows $(\bar{c}_i^\nu . \bar{c}_j^{n-\nu}) = a(c_i^\nu . c_j^{n-\nu})$ and generally

$$(\bar{c}^\nu . \bar{c}^{n-\nu}) = a(c^\nu . c^{n-\nu}) \tag{5'}$$

for arbitrary cycles c^ν, $c^{n-\nu}$. Therefore, we have proved

THEOREM V. *When M undergoes a one-valued transformation of degree $\neq 0$, all Kronecker-indices of cycle-pairs are multiplied by the degree.*

As a corollary of this theorem we have the fact that *the property of a pair of cycles c^ν, $c^{n-\nu}$ to intersect or not to intersect one another in the algebraic sense is not only invariant under homeomorphisms of M^n but under all transformations of M^n into itself, which have a degree $\neq 0$.*

We may point out that the assumption $a \neq 0$ which we have used in deriving (3) and, therefore, in proving V is necessary; for a simple transformation of a surface of genus 2 shows that the statements of theorem V and of its corollary are not correct in the case $a = 0$.

6. *The Method of the Product Manifold*, introduced by Lefschetz, will be used in proving the contragredience formula (2). It may be described as follows:[5]

If x and ξ are points of M^n and μ^n, respectively, and $X = x \times \xi$ the point of the product $M^n \times \mu^n$ which represents the pair x, ξ, then we write

$$x = P(X), \quad \xi = \Pi(X) \quad [X = x \times \xi] \tag{6}$$

and call P and Π the "projections" on M^n and μ^n. We consider an n-cycle Γ^n in $M^n \times \mu^n$ which generally is singular and say that x and ξ are corresponding with respect to Γ^n if there exists an X on Γ^n, so that (6) hold. If we interpret ξ as the image of x then this correspondence defines a "transformation" T of M^n into μ^n, which may be symbolically expressed by

$$T(x) = \Pi P^{-1}(x) \tag{7a}$$

and similarly a transformation T^{-1} of μ^n into M^n, the "inverse" of T:

$$T^{-1}(\xi) = P\Pi^{-1}(\xi). \tag{7b}$$

T and T^{-1} are both, in general, multiply valued.

The following construction gives the image $T(c^\nu)$ of a ν-cycle c^ν of M^n: the "cylinder" $c^\nu \times \mu^n$ erected on c^ν in $M^n \times \mu^n$ intersects Γ^n in a ν-cycle $\Gamma^n . c^\nu \times \mu^n$ and the projection of this cycle on μ^n is the image of c^ν:

$$T(c^\nu) = \Pi(\Gamma^n . c^\nu \times \mu^n). \tag{8a}$$

Similarly, we determine

$$T^{-1}(\gamma^\nu) = P(\Gamma^n . M^n \times \gamma^\nu) \tag{8b}$$

for each cycle γ^ν of μ^n.

Concerning fundamental sets of cycles in $M^n \times \mu^n$ we have the theorem[6] that, if c_i^λ, γ_j^λ ($\lambda = 0, 1, \ldots, n$; $i = 1, 2, \ldots, p_\lambda$; $j = 1, 2, \ldots, \pi_\lambda$), are fundamental sets in M^n and μ^n, the set of all products $c_i^\lambda \times \gamma_j^{\nu-\lambda}$ ($\lambda = 0, 1, \ldots, \nu$) forms a fundamental set of ν-cycles in $M^n \times \mu^n$. The projection of $\Delta^\nu = c_i^\lambda \times \gamma_j^{\nu-\lambda}$ on M^n

$$P(\Delta^\nu) = P(c_i^\lambda \times \gamma_j^{\nu-\lambda}) = c_i^\lambda$$

is, considered as ν-cycle in M^n, ≈ 0, if $\lambda < \nu$; likewise is

$$\Pi(\Delta^\nu) = \Pi(c_i^\lambda \times \gamma_j^{\nu-\lambda}) = \gamma_j^{\nu-\lambda}.$$

≈ 0 on μ^n, if $\lambda > 0$. Thus, if we have any ν-cycle

$$\Gamma^\nu \approx \sum_{\lambda=0}^{\nu} \sum_{i,j} \eta_{ij}^{\lambda} . c_i^{\lambda} \times \gamma_j^{\nu-\lambda} \quad (i = 1, 2, \ldots, p_\lambda;\ j = 1, 2, \ldots, \pi_\lambda) \quad (9)$$

then

$$P(\Gamma^\nu) \approx \sum_{i=1}^{p_\nu} \eta_{i1}^{\nu} c_i^{\nu} \quad [\text{on } M^n] \quad (9a)$$

$$\Pi(\Gamma^\nu) \approx \sum_{j=1}^{\pi_\nu} \eta_{ij}^{0} \gamma_j^{\nu} \quad [\text{on } \mu^n]. \quad (9b)$$

Now let the cycle Γ^n which defines the transformation T be

$$\Gamma^n \approx \sum_{\lambda=0}^{n} \sum_{i,j} \epsilon_{ij}^{\lambda} . c_i^{\lambda} \times \gamma_j^{n-\lambda}. \quad (10)$$

In calculating $T(c_k^\nu)$ and $T^{-1}(\gamma_k^\nu)$ by (10), (8a), (8b) we shall make use of[7]

$$(c^p \times \gamma^r . c^q \times \gamma^s) \approx (-1)^{(n-p)(n-s)} (c^p . c^q) \times (\gamma^r . \gamma^s) \ [\text{on } M^n \times \mu^n]. \quad (11)$$

From (10), (11) and $(\gamma^{n-\lambda} . \mu^n) = \gamma^{n-\lambda}$ we find

$$\Gamma^n . c_k^\nu \times \mu^n \approx \sum_{\lambda=0}^{n} \sum_{i,j} \epsilon_{ij}^{\lambda} (c_i^{\lambda} \times \gamma_j^{n-\lambda} . c_k^{\gamma} \times \mu^n)$$

$$\approx \sum_{\lambda=0}^{n} \sum_{i,j} \epsilon_{ij}^{\lambda} (c_i^{\lambda} . c_k^{\nu}) \times \gamma_j^{n-\lambda}.$$

In view of $(c_i^\lambda . c_k^\nu) = 0$, if $\lambda + \nu < n$, and of (9b), for $\Pi(\Gamma^n . c_k^\nu \times \mu^n)$ the only terms that are essential are those in which $(\gamma_j^{n-\lambda} . \mu^n)$ is ν-dimensional, i.e. $\lambda = n - \nu$; hence,

$$T(c_k^\nu) = \Pi(\Gamma^n . c_k^\nu \times \mu^n) \approx \sum_{i,j} \epsilon_{ij}^{n-\nu} . (c_i^{n-\nu} . c_k^\nu) . \gamma_j^\nu. \quad (12a)$$

Therefore, if

$$T(c_k^\nu) \approx \sum_{j=1}^{\pi_\nu} \alpha_{kj}^\nu \gamma_j^\nu, \quad (13a)$$

then we get from (12a)

$$\alpha_{kj}^\nu = \sum_{i=1}^{p_\nu} e_{ij}^{n-\nu} (c_i^{n-\nu} . c_k^\nu)$$

or in terms of matrices, with $\epsilon_{n-\nu} = \| e_{ij}^{n-\nu} \|$, $A_\nu = \| \alpha_{kj}^\nu \|$

$$A_\nu = L'_{n-\nu} \epsilon_{n-\nu} = (-1)^{\nu(n+1)} L_\nu \epsilon_{n-\nu}.[8] \quad (14a)$$

Similarly, let be for T^{-1}

$$T^{-1}(\gamma_k^\nu) \approx \sum_{i=1}^{p_\nu} \beta_{ki}^\nu c_i^\nu.$$

Then by means of (10), (11) and in view of $(c^\lambda . M^n) = c^\lambda$:

$$\Gamma^n . M^n \times \gamma_k^\nu \approx \sum_{\lambda=0}^{n} \sum_{ij} \epsilon_{ij}^\lambda . c_i^\lambda \times \gamma_j^{n-\lambda} . M^n \times \gamma_k^\nu$$

$$\approx \sum_{\lambda=0}^{n} \sum_{ij} \epsilon_{ij}^\lambda (-1)^{(n-\lambda)(n-\nu)} c_i^\lambda \times (\gamma_j^{n-\lambda} . \gamma_k^\nu).$$

According to $(\gamma^{n-\lambda} . \gamma^\nu) = 0$ for $\lambda > \nu$ and to (9a) we are only interested in those terms on the right hand where c_i^λ is of dimensionality ν, i.e., $\lambda = \nu$; hence,

$$T^{-1}(\gamma_k^\nu) = P\,(\Gamma^n . M^n \times \gamma_k^\nu) \approx \sum_{i,j} \epsilon_{ij}^\nu (-1)^{n-\nu} (\gamma_j^{n-\nu} . \gamma_k^\nu) c_i^\nu \quad (12b)$$

and thus from (13b)

$$B_{ki}^\nu = (-1)^{n-\nu} \sum_{j=1}^{\pi_\nu} \epsilon_{ij}^\nu (\gamma_j^{n-\nu} . \gamma_k^\nu),$$

i.e., if $B_\nu = \| \beta_{ki}^\nu \|$

$$B_\nu = (-1)^{n-\nu} \Lambda'_{n-\nu} \epsilon'_\nu. \quad (14b)$$

Now one sees that between the matrices A and B which define the cycle transformations under T and T^{-1} there must hold a certain relation: On replacing ν by $n - \nu$ in (14a) we find $\epsilon_\nu = L_\nu'^{-1} A_{n-\nu}$, and since $L_{n-\nu} = (-1)^{\nu(n-1)} L'_\nu$,

$$\epsilon'_\nu = (-1)^{\nu(n+1)} A'_{n-\nu} L'^{-1}_{n-\nu}.$$

Therefore, from (14b)

$$B_\nu = (-1)^{n(\nu+1)} \Lambda'_{n-\nu} A'_{n-\nu} L'^{-1}_{n-\nu} = (-1)^{n(\nu+1)} (L^{-1}_{n-\nu} A_{n-\nu} \Lambda_{n-\nu})'. \quad (15)$$

Let now a cycle γ_k^ν of μ^n be transformed into $c^\nu = T^{-1}(\gamma_k^\nu)$ and then back into $T(c^\nu) = TT^{-1}(\gamma_k^\nu) = \bar{\gamma}_k^\nu$. We have

$$TT^{-1}(\gamma_k^\nu) = \bar{\gamma}_k^\nu \approx \sum_{j=1}^{\pi_\nu} u_{kj}^\nu \gamma_j^\nu$$

with a square matrix $U_\nu = \| u_{kj}^\nu \|$. Then the homologies

$$T^{-1}(\gamma_k^\nu) = c^\nu \approx \sum_{i=1}^{p_\nu} \beta_{ki}^\nu c_i^\nu$$

$$T(c^\nu) = \bar{\gamma}_k^\nu \approx \sum_i \beta_{ki}^\nu T(c_i^\nu) = \sum_{i,j} \beta_{ki}^\nu \alpha_{ij}^\nu \gamma_j^\nu$$

show that

$$B_\nu A_\nu = U_\nu.$$

Hence the matrices U_ν belonging to the transformation TT^{-1} of μ^n into

itself are according to (15) expressible by the matrices A in the following manner:

$$U_\nu = (-1)^{n(\nu+1)} \cdot (L_{n-\nu}^{-1} A_{n-\nu} \Lambda_{n-\nu})' A_\nu. \tag{16}$$

7. *Proof of the Contragredience Formula (2).*—Up to this point we did not separate one-valued and multiply-valued transformations T. The introduction of TT^{-1} marks the place where they naturally part. If we go from γ^ν to $T^{-1}(\gamma^\nu) = c^\nu$ and from c^ν back to $T(c^\nu) = TT^{-1}(\gamma^\nu) = \overline{\gamma}^\nu$, then in the general case we do not return to γ^ν. For, although the cylinder $c^\nu \times \mu^n$ has in common with Γ^n the cycle $\Gamma^n . M^n \times \gamma^\nu$, of which the projection on μ^n is γ^ν, this cycle is not the whole intersection $\Gamma^n . c^\nu \times \mu^n$ and, therefore, the projection $\Pi(\Gamma^n . c^\nu \times \mu^n)$ is different from γ^ν. When, on the contrary, T is *one-valued*, then for each point x of M^n the product $x \times \mu^n$ has only the point $X = x \times T(x)$ in common with Γ^n; therefore, for each cycle Δ^ν of Γ^n the cylinder $P(\Delta^\nu) \times \mu^n$ intersects Γ^n only in the points of Δ^ν; hence,

$$\Gamma^n . P(\Delta^\nu) \times \mu^n = \Delta^\nu. \tag{17}$$

If we take $\Delta^\nu = \Gamma^n . M^n \times \gamma^\nu$, then from (8*b*) and (17) follows

$$\Gamma^n . T^{-1}(\gamma^\nu) \times \mu^n = \Gamma^n . M^n \times \gamma^\nu \tag{18}$$

and from (8*a*)

$$TT^{-1}(\gamma^\nu) = \Pi(\Gamma^n . M^n \times \gamma^\nu) \tag{19}$$

for each cycle γ^ν of μ^n. The right hand of this equation may be determined by the method which has yielded (12*a*); from (10), (11) and $(c^\lambda . M^n) = c^\lambda$ follows

$$\Gamma^n . M^n \times \gamma^\nu \approx \sum_{\lambda=0}^{n} \sum_{ij} \epsilon_{ij}^\lambda (-1)^{(n-\lambda)(n-\nu)} c_i^\lambda \times (\gamma_j^{\lambda-\lambda} . \gamma^\nu).$$

For the projection Π we may omit all terms $(\gamma_j^{n-\mu} . \gamma^\nu)$ with dimensionality $\neq \nu$, i.e., with $\lambda \neq 0$; hence,

$$\Pi(\Gamma^n . M^n \times \gamma^\nu) \approx \sum_{i,j} \epsilon_{ij}^0 (-1)^{n(n-\nu)} \gamma^\nu.$$

But because $p_0 = \pi_0 = 1$, the matrix $\| \underset{ij}{\epsilon_0} \|$ has only one element ϵ_0; from (14*a*) follows $\epsilon_0 = a_n$, where a_n is the only element of the matrix A_n and this element is by definition (cf. Nr. 1), the *degree* a of T. Hence,

$$\Pi(\Gamma^n . M^n \times \gamma^\nu) \approx (-1)^{n(\nu+1)} a \gamma^\nu \tag{20a}$$

and from (19)

$$TT^{-1}(\gamma^\nu) \approx (-1)^{n(\nu+1)} a \gamma^\nu. \tag{20b}$$

Therefore, the matrix U_ν defined in Nr. 6 has been determined under the assumption that T is one-valued:

$$U_\nu = (-1)^{n(\nu+1)} a E_{\pi_\nu} \tag{21}$$

where E_{π_ν} is the matrix unity. From (21) and (16) formula (2) follows immediately.

[1] Cf. S. Lefschetz, (*a*) *Trans. Am. Math. Soc.*, **28**, pp. 1–49; (*b*) *Trans. Am. Math. Soc.*, **29**, pp. 429–462; particularly p. 32 of (*a*). The sign "$\approx$" introduced in (*a*) means "$\sim$" mod. zero-divisors. The fundamental sets of this paper are all with respect to the operation $\approx$.

[2] L. E. Y. Brouwer, *Mathem. Annalen*, **71**, pp. 92–115.

[3] O. Veblen, *Trans. Am. Math. Soc.*, **25**, pp. 540–550.

[4] (*a*), formula 1.1; (*b*), formula 10.5.

[5] A great deal of Nr. 6 is only a summarizing report on facts which are included in the papers of Lefschetz, quoted above.

[6] Lefschetz, (*a*) No. 52.

[7] Lefschetz, (*a*) No. 55. The proof of this formula, not explicitly given there, can be obtained easily by the same considerations as for the formulas of (*a*) Nos. 53, 54.

[8] Lefschetz, (*b*) 9.2.

HARMONY AS A PRINCIPLE OF MATHEMATICAL DEVELOPMENT

By G. A. Miller

Department of Mathematics, University of Illinois

Communicated January 27, 1928

In the preface of volume 6 of the "Collected Works" of Sophus Lie the editor remarked that these works constitute an artistic contribution which is entirely comparable with that of Beethoven. In both cases a few apparently trivial motives dominated the entire creation. In the case of Lie it was mainly the explicit use of the concept of group in domains where its dominance had not been explicitly recognized before his time. This concept was practically refused by the mathematical builders up to the beginning of the nineteenth century but became during this century "the head stone of the corner" largely through the work of Sophus Lie. While the work of Beethoven has reached and probably will continue to reach a much larger number of people than that of Lie, it is questionable whether those to whom Lie's work is actually revealed receive less inspiration therefrom or are less impressed by the marvelous new harmonies which it introduces into a wide range of mathematical developments. As Sophus Lie was a Foreign Associate of this Academy and his work

has greatly influenced the development of mathematics in our country it seems appropriate to note here this tribute to his memory.

A. Speiser recently directed attention to the importance of the study of the symmetries of ancient ornaments with a view to an understanding of their influence on the development of mathematics. In particular, he expressed the view that an investigation of the geometric content of Egyptian and Arabian ornaments constitutes one of the most beautiful chapters of the history of mathematics which remains to be written, and that the development of higher mathematics began about 1000 years earlier than is now commonly assumed, having been inspired by an interest in regular figures on the part of the ancient Egyptians as early as 1500 B. C. Evidence of his faith in the fruitful contact between mathematics and ancient ornaments is exhibited by the remark that an intensive study of the noted work entitled "The Grammar of Ornament" by Owen Jones, **1856,** may be urgently recommended for all who are interested in the theory of groups.[1] This new point of view is emphasized here because it tends to widen the horizon of the specialist in this field and to unite scientific efforts along lines which are commonly regarded as widely apart.

Our main present object is to exhibit the principle of harmony as a useful principle in the study of the history of mathematics. The difficulties of this study are increasing rapidly with the growth of our subject and hence it is increasingly desirable to find general coördinating and clarifying principles. It is well known that in the development of the number concept the principle of permanence of form, which has much in common with the principle of harmony, played a very fundamental rôle. The art of combining with a minimum of blind formulas a maximum of seeing thoughts, which Minkowski characterized as the true Dirichlet Principle,[2] is largely due to the principle of harmony in mathematics. H. Poincaré,[3] noted that without mathematics "we should have been ignorant forever of the internal harmony of the world, which is the only true objective reality." It is only natural that the language which is so powerful for discovering the harmonies of nature should be dominated by the same fundamental principle. This is reflected in the fact that the adjective *harmonic* is used in connection with a very large number of mathematical terms.

The ancient Greeks introduced a very fundamental note of discord into the development of mathematics when they assumed that there is no (1, 1) correspondence between the real numbers and the points on a straight line. This discord was largely removed by Newton when he defined a number as the ratio of one line segment to another. Here, as well as in various other parts of mathematics, perfect harmony was finally established by means of an explicit postulate. Since the times of the ancient Greeks mathematicians have never been in complete accord as regards some of

the fundamental principles. Fortunately, there have always been also large domains of mathematics which have been free from discord.

Among the many fundamental developments in elementary mathematics which seem to have been largely influenced by the principle of harmony we may note here the slow but continued progress towards supplanting the proportion by the equation. It is well known that the proportion played a much more prominent rôle in Greek mathematics than the equation, and that even in the eighteenth century some of the textbooks considered the theory of proportion at undue length. It is clear that the equation naturally implies a higher degree of knowledge than the proportion, and hence it implies greater intellectual harmony. The long struggle towards supplanting the proportion by the equation is not even now at an end, but progress in this direction has been very marked. A similar movement towards harmony may be seen in the prolonged effort towards replacing the definitions of the trigometric functions, depending both on the angle and the length of the line segments involved, by definitions depending only on the former. Fortunately, in this case the victory towards greater harmony in the formulas is now complete.

Developments in mathematics have frequently been made more harmonious by the introduction of ideal elements. As a conspicuous instance we may cite the use of imaginary numbers from about the middle of the sixteenth century to the close of the eighteenth, when the use of these numbers began to be placed on the basis of reals by the work of Caspar Wessell and others. Early in the seventeenth century Kepler noted that parallel lines may be regarded as having a point at infinity in common, and this use of the point at infinity later became general through the work of Desargnes and others. As a much later instance of the introduction of ideal elements we may refer to the ideals introduced by Kummer for the study of algebraic numbers. It is true that the introduction of such ideal elements can also be explained as having been actuated by the principle of generalizations, but the more general principle of harmony is clearly also involved therein. As harmony in music makes an almost universal appeal to the human ear so harmony in scientific matters makes an almost universal appeal to the human intellect, and many of the developments in mathematics can be more fully understood if this principle is duly emphasized.

It is commonly recognized that the mathematical results are more interdependent than those of any other subject and hence *the history of mathematics is naturally more scientific than any other history.* The principle of harmony is not only a source of inspiration for the investigator but also a guide as regards the novelty of the results obtained. Moreover, just as an increase in the accuracy of measurement usually discloses greater discrepancies, so an increase in accuracy of knowledge usually discloses

new elements of discord. Hence, the principle of harmony tends to become more effective with the advance of knowledge and to dominate the energies of the investigator more powerfully as he attacks problems of a more fundamental nature.

[1] Speiser, A., "Die Theorie der Gruppen von Endlicher Ordung," **1927,** p. 77.

[2] Klein, F., "Vorlesungen über die Enlivicklung der Mathematik im 19. Jahrhundert," **1926,** p. 97.

[3] Poincaré, H., "La Valeur de la Science," **1908,** p. 7.

NOTE ON THE BEHAVIOR OF CERTAIN POWER SERIES ON THE CIRCLE OF CONVERGENCE WITH APPLICATION TO A PROBLEM OF CARLEMAN

BY EINAR HILLE

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated February 11, 1928

1. The first example of a continuous periodic function whose associated Fourier series $\sum_{-\infty}^{+\infty} c_\nu e^{i\nu x}$ is such that $\sum |c_\nu|^\rho$ converges for no $\rho < 2$ was given by Carleman.[1] Landau,[2] who simplified the example, called attention to a power series studied by Fabry[3] and Hardy[4] which series is continuous on the unit circle, the series $\sum |c_\nu|^{2-\delta}$ being divergent for a fixed but arbitrarily small $\delta > 0$. Hardy had to use a powerful machinery in his study of the singularities of this function; if one is satisfied, however, with determining merely the properties of convergence of the series on the unit-circle, a simpler argument can be used which applies to a much larger class of series. The method which I use for this purpose is based on some recent applications of Weyl's ideas regarding equi-distributed point-sets to number-theoretic questions due to van der Corput.[5] A fairly simple solution of Carleman's problem is obtained in this manner.

2. We shall study a class of power series

$$\sum_{n_0}^{\infty} f(n) \exp[2\pi i a(n)]\, z^n \tag{1}$$

where $f(n)$ and $a(n)$ are subjected to one of the following three sets of conditions:

I. (*i*) $a(u)$ is a real differentiable function when $u \geqq n_0$ and $\lim a(u) = +\infty$;

(*ii*) $a'(u)$ is positive, never increasing and $\lim a'(u) = 0$;

(*iii*) $\sum |f(n) - f(n+1)|$ converges and $\lim f(n) = 0$.

II. (*i*) and (*ii*) as in I;
(*iii*) $\sum |f(n) - f(n+1)|\,[a'(n)]^{-1}$ converges and $\lim f(n)[a'(n)]^{-1} = 0$.
III. (*i*) and (*ii*) as in I;
(*iii*) $a''(u)$ exists and is never decreasing but remains <0;
(*iv*) $\sum |f(n) - f(n+1)|\,|a''(n)|^{-1/2}$ converges and $\lim f(n)\cdot|a''(n)|^{-1/2} = 0$.

We shall say that case N occurs when our series satisfies the set of conditions N.

THEOREM. *In case I the series* (1) *converges when* $|z| \leqq 1$, $z \neq +1$. *The convergence is uniform when* $|z| \leqq 1$, $|1 - z| \geqq \epsilon$. *In case II the series converges when* $|z| \leqq 1$ *and the convergence is uniform when* $|z| \leqq 1$, $\arg(1 - z) \leqq \pi/2 - \epsilon$. *In case III the series converges uniformly for* $|z| \leqq 1$.

3. Writing

$$z = \exp(2\pi i t),\quad P_{m,n}(t) = \sum_{k=m}^{n} f(k)\exp\{2\pi i[a(k) + tk]\},$$

$$S_{m,n}(t) = \sum_{k=m}^{n} \exp\{2\pi i[a(k) + tk]\},$$

we get

$$P_{m,n}(t) = \sum_{k=m}^{n} [f(k) - f(k+1)]\,S_{m,k}(t) + f(n)\,S_{m,n}(t)$$

where $m \geqq n_0$ is to be suitably chosen. Thus in order to prove the theorem we have merely to show the existence of constants K_ϵ, B and C, independent of k, such that

$$|S_{m,k}(t)| \leqq K_\epsilon \quad \text{, when } 0 < \epsilon \leqq t \leqq 1 - \epsilon, \tag{2}$$

$$|S_{m,k}(t)| \leqq B[a'(k)]^{-1} \quad \text{, when } 0 \leqq t \leqq 1/2, \tag{3}$$

$$|S_{m,k}(t)| \leqq C|a''(k)|^{-1/2}, \text{ when } 0 \leqq t \leqq 1, \tag{4}$$

in cases I, II and III, respectively.[6]

4. The proof of these inequalities is based on the following three lemmas:

LEMMA 1. Let $F(u)$ be real and differentiable in the interval $m \leqq u \leqq n$, let $F'(u)$ be monotone and $|F'(u)| \leqq 1 - \delta$, $\delta > 0$ fixed. Then there exists a constant A_δ such that

$$\left|\sum_{m}^{n} \exp[2\pi i\,F(\nu)] - \int_{m}^{n} \exp[2\pi i\,F(u)]du\right| < A_\delta.$$

This lemma is due to van der Corput[7] who considered the case $\delta = 1/2$. It follows from his proof in this case that

$$A_\delta < 1 + \frac{1-\delta}{2}\sum_{\nu=1}^{\infty}\frac{1}{\nu(\nu - 1 + \delta)}.$$

LEMMA 2. Let $F'(u)$ be monotone decreasing but remain >0, then

$$\left| \int_m^n \exp[2\pi i\, F(u)]du \right| \leqq \frac{1+\sqrt{2}}{2\pi} [F'(n)]^{-1}.$$

This follows from the identity

$$\int_m^n \exp[2\pi i\, F(u)]du = (2\pi i)^{-1} \int_m^n [F'(u)]^{-1} d \exp[2\pi i\, F(u)]$$

by applying Bonnet's form of the second mean value theorem.

LEMMA 3. Let $F'(u)$ be monotone decreasing and suppose that $F''(u)$ exists and is monotone increasing but remains <0. Then

$$\left| \int_m^n \exp[2\pi i\, F(u)]du \right| \leqq {}^5/_2[-F''(u)]^{-1/2}.$$

This lemma, save for the value of the numerical constant, is a special case of van der Corput's Hilfsatz 2 (loc. cit., p. 62). Its proof follows readily from lemma 2.

5. We can now set

$$F(u) = a(u) + t\,u.$$

Assuming $-1+\epsilon \leqq t \leqq 1-\epsilon$, we can find an $m \geqq n_0$ and a $\delta > 0$ such that

$$|F'(u)| = |a'(u) + t| \leqq 1 - \delta, \text{ when } u \geqq m.$$

Hence, we get by lemma 1

$$|S_{m,n}(t)| \leqq A_\delta + \left| \int_m^n \exp[2\pi i\, F(u)]du \right|,$$

and by lemma 2 the integral is less than

$${}^2/_5[a'(n) + t]^{-1}$$

provided $0 \leqq t \leqq 1 - \epsilon$. This estimate is sufficient to prove formulas (2) and (3). In case III we can apply lemma 3 for the estimating of the integral, obtaining

$$|S_{m,n}(t)| \leqq A_\delta + {}^5/_2[-a''(n)]^{-1/2}$$

which proves formula (4).

6. As an example to illustrate our theorem we can choose

$$a(n) = n^\alpha,\ f(n) = n^{-\beta},\ 0 < \alpha < 1,\ 0 < \beta.$$

The corresponding series

$$P(z, \alpha, \beta) = \sum_{n=1}^{\infty} n^{-\beta} \exp(2\pi i\, n^\alpha)\, z^n \qquad (5)$$

is the series of Fabry and Hardy mentioned in §1. Applying our theorem, we find that the series converges everywhere on the unit circle when

$\alpha + \beta > 1$, and uniformly if $\alpha/2 + \beta > 1$. These are the precise limits found by Hardy.[8] Thus the function $P(z, \alpha, \beta)$ is continuous on the unit circle when $\beta > 1 - \alpha/2$. This quantity can be made as close as we please to $^1/_2$ by choosing α near to 1. On the other hand, the series $\sum_{n=1}^{\infty} (n^{-\beta})^{\rho}$ converges only for $\rho > 1/\beta$ which can be made as close as we please to 2. This is the essence of Landau's observation.

As a second example we take

$$a(n) = n(\log n)^{-\alpha},\ f(n) = n^{-1/2}(\log n)^{-\beta},\ 0 < \alpha,\ 0 < \beta.$$

The corresponding series

$$P_{\alpha,\beta}(z) = \sum_{n=2}^{\infty} n^{-1/2}(\log n)^{-\beta} \exp[2\pi i\, n(\log n)^{-\alpha}]\, z^n \qquad (6)$$

converges everywhere on the unit circle and the convergence is uniform whenever $\beta > {}^1/_2(\alpha + 3)$. The series $\sum [f(n)]^{\rho}$ is evidently never convergent for any $\rho < 2$. *Thus the function $P_{\alpha,\beta}(z)$ is a solution of Carleman's problem whenever $\beta > {}^1/_2(\alpha + 3)$.* The functions $P(z, \alpha, \beta)$ and $P_{\alpha,\beta}(z)$ are both infinitely many-valued and their only singularities are 0, 1 and ∞.[9]

A more general solution of Carleman's problem can be found as follows. Let $a(n)$ and $f(n)$ satisfy conditions III and the additional restriction that $\sum [f(n)]^{\rho}$ shall be convergent for $\rho \geqq 2$ and divergent for $\rho < 2$. The corresponding series (1) will be a solution of Carleman's problem.

[1] T. Carleman, *Acta Math.*, **41**, 1918 (377–389).

[2] E. Landau, *Math. Zeitschrift*, **5**, 1919 (147–153). See also O. Szász, *Ibid.*, **8**, 1920 (222–236) and T. H. Gronwall, *Bull. Amer. Math. Soc.*, **27**, 1921 (320–321).

[3] E. Fabry, *Acta Math.*, **36**, 1913 (69–104).

[4] G. H. Hardy, *Quart. Journ.*, **44**, 1913 (147–160). Cf. G. H. Hardy and J. E. Littlewood, these PROCEEDINGS, **2**, 1916 (583–586).

[5] J. G. van der Corput, *Math. Annalen*, **84**, 1921 (53–79). Extensions are to be found *Ibid.*, **87**, 1922 (39–65) and **89**, 1923 (215–254). Cf. Landau, *Acta Math.*, **48**, 1926 (217–263).

[6] The uniform convergence of the series in the regions stated in the theorem follows readily from the uniform convergence on the various arcs of the unit circle implied by the estimates (2)–(4).

[7] First paper quoted in note 5, p. 58, Satz 1. A more general theorem is to be found on p. 40 of the second paper.

[8] Loc. cit., p. 157. That the series (5) diverges when $\alpha + \beta < 1$ has been shown by Hardy in *Proc. London Math. Soc.* (2) **9**, 1911 (126–144).

[9] Proved by Hardy for series (5) who gives a thorough discussion of the nature of the singularity at $z = +1$. The location and possibly also the nature of the singularities can be determined for both series with the aid of a theorem due to E. Le Roy and extended by E. Lindelöf. See the latter's *Calcul des residus*, chapter V, especially pp. 109 and 124. This method can be used for an attack on the problem proposed by Hardy on p. 160 of his paper; it cannot give a complete solution of the problem, but such a solution is scarcely to be expected from the nature of the problem.

INTRA-SEASONAL CYCLES OF GROWTH*

BY H. S. REED

UNIVERSITY OF CALIFORNIA, COLLEGE OF AGRICULTURE

Communicated February 8, 1928

The growth process of organisms is often rhythmic, resulting in the production of well-marked growth cycles. If the growth cycles overlap to any great extent their individual forms are not readily distinguishable, but their presence is generally evident, at least in the more complex organisms. Their study gains interest from the fact that certain physical chemists[3] have initiated a noteworthy study of periodic phenomena in heterogeneous systems. The complicated processes of organic growth are admittedly more difficult to study than the simpler inorganic systems of the chemist's beaker, yet there are resemblances which cannot be disregarded. Annual fluctuations in the growth rate of plants are well known, but the intra-seasonal cycles are less prominent. In the cases studied the oscillations of the growth rate diminish from the beginning to the end of the annual growing season.[5,6,8] One might infer that some agent (in either the environment or the heredity of the organism) damps the oscillations. Evidence has been presented[7,9] to show that these cycles of growth are actually related to other physiological conditions, e.g., the growth rate of the shoots showed a strong negative correlation with their sap concentration.

I have recently undertaken further studies on the problem of intra-seasonal growth cycles, utilizing measurements of the growth of 94 shoots of lemon trees (*Citrus limonia* Osbeck), growing in a grove at Riverside, California.

Dr. F. F. Halma kindly assisted in taking the measurements during the growing season, which extended from May 1st to October 23rd. The three distinct cycles of growth, each of which ran for about eight weeks, are shown in figure 1. The shoots made approximately the same amount of new growth in the first and second cycles, but fell short of that amount in the third. The curve of seasonal growth of the population is shown by a composite of the three sigmoid curves and shows the characteristic periodic fluctuations.

The rate of growth in each cycle is expressed by the differential equation

$$\frac{dx}{dt} = k\,x(A - x).$$

The differential equation may also be written

$$\frac{dx}{dt} = k_1(C)$$

where the activity of the catalyst (C) is proportional to $x(A - x)$, increasing from a minimum to a maximum (Fig. 4) and again approaching zero at the end of the growth cycle.

The growth process of each cycle conforms to a curve expressed by the integral equation

$$\log \frac{x}{A - x} = K(t - t_1)$$

in which x = length of shoots at time t; A = final length; t_1 = time at which $x = \frac{A}{2}$, i.e., when the cycle of growth is half completed; k = a specific constant of the process, and $K = Ak$.

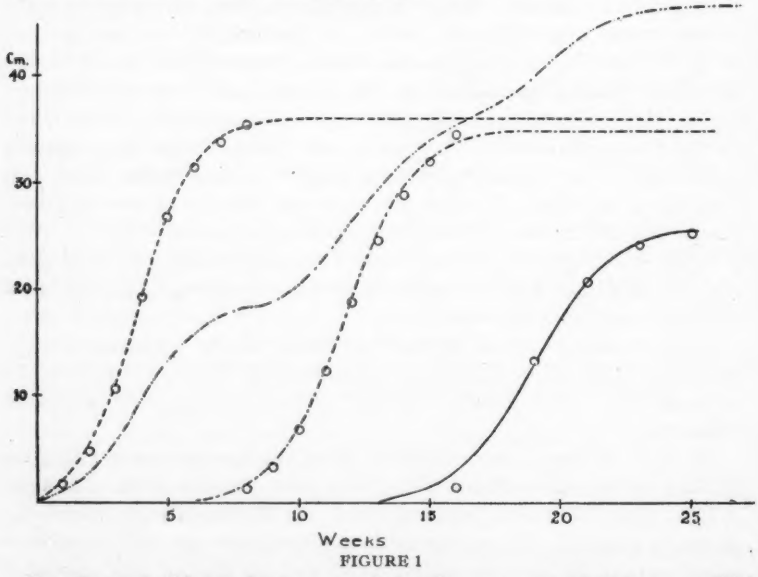


FIGURE 1

Growth curves of lemon shoots, - - - - first cycle, — - — second cycle, ——— third cycle, —··—··— summation of growth of the individual cycles plotted on a scale of ordinates half as great as that employed for the separate cycles. Circles represent observed values.

The equations for the three curves are

$$\text{Cycle I,} \quad \log \frac{x}{36 - x} = 0.411\ (t - 3.85)$$

$$\text{Cycle II,} \quad \log \frac{x - 34}{69 - x} = 0.336\ (t - 11.75)$$

$$\text{Cycle III,} \quad \log \frac{x - 67}{93 - x} = 0.311\ (t - 19.00).$$

The values of A and of t_1 were obtained by approximation, those of K are the averages of the individual constants for the respective cycles. The values of x calculated from these equations are given in table 1. The experimentally obtained values of K were larger at the outset of each cycle and declined, at first rapidly, then more slowly as the cycle progressed. The values of log K plotted against values of $\frac{x}{A}$, the per cent of growth attained at any given interval, give approximately parallel straight lines (Fig. 2). This results in a slight asymmetry of the growth curve, but it seems advisable to leave this asymmetry uncorrected, although equations giving a closer fitting curve (10) may be used.

TABLE 1

LEMON SHOOTS. OBSERVED AND CALCULATED VALUES OF LENGTH

Cycle I				Cycle II				Cycle III			
t	x (OBSERVED)	K	x (CALCULATED)	t	$x - 34$ (OBSERVED)	K	$x - 34$ (CALCULATED)	t	$x - 67$ (OBSERVED)	K	$x - 67$ (CALCULATED)
WEEKS	CM.		CM.	WEEKS	CM.		CM.	WEEKS	CM.		CM.
0	...	...	0.09	8	1.19	0.388	1.82	16	1.49	0.406	2.73
1	1.69	0.459	2.27	9	3.26	0.360	3.74	17	5.13	0.305	5.02
2	4.81	0.438	5.32	10	6.89	0.348	7.17	19	13.30	...	13.00
3	10.59	0.447	11.11	11	12.35	0.351	12.56	21	20.64	0.293	20.98
4	19.26	0.406	19.25	12	18.72	0.244	19.18	23	24.28	0.288	24.60
5	26.62	0.395	26.91	13	24.53	0.296	25.34	25	25.32	0.262	25.66
6	31.33	0.385	31.83	14	28.90	0.301	29.78				
7	33.72	0.372	34.25	15	31.96	0.314	32.37				
8	35.19	0.388	35.30	16	34.49	0.427	33.74				

Since the equations indicate that growth proceeded like a series of autocatalytic reactions in which the activity of the catalyst (C) is proportional to $x(A - x)$, it is logical to adopt the idea of an autocatalyst of growth as a working hypothesis. The values of K then represent the activity of some growth-promoting agency in the system, without which none of the external substrates would be converted into organized substance. The general slope of each growth curve represents the activity of the autocatalyst of that particular cycle. The mean values of K were progressively smaller in successive cycles (table 2) and indicate that the activity of the autocatalyst had a downward trend from the beginning to the end of the growing season (Fig. 3). The individual values of K at the outset of the second and third cycles rose, however, to a point nearly level with that at the beginning of the first cycle. The value of K determines the slope of the growth curve (Fig. 1) and may be taken as a constant of the growth process for each cycle. The value of $k\left(=\frac{K}{A}\right)$ is,

however, independent of A and represents the actual constant of growth in any cycle. The values of log k in the different cycles show a similar parallelism.

The distinct cycles and their sequence is well shown by the derivatives of the growth curves (Fig. 4). The cycles, although distinct, overlap to a slight extent, owing in most instances to the fact that the growth of various individual shoots was not absolutely synchronous. A few shoots continued after the majority of the population had ceased growing, causing the skewness which became pronounced in the third cycle.

TABLE 2

CONSTANTS OF THE CURVES OF GROWTH OF THREE INTRA-SEASONAL CYCLES OF LEMON SHOOTS FOR THE GENERAL EQUATION

$$\text{Log}\frac{x}{A-x} = K(t-t_1)$$

	CYCLE		
ALL SHOOTS	I	II	III
A	36 cm.	35 cm.	26 cm.
K	0.411 ± 0.0072	0.336 ± 0.012	0.311 ± 0.0222
k	0.0114	0.0096	0.0120
Root-mean-square deviation of calculated values of x	0.43 cm.	0.59 cm.	0.57 cm.
"ONE-CYCLE" SHOOTS			
A	29 cm.		
K	0.473 ± 0.021		
k	0.0163		
Root-mean-square deviation of calculated values of x	0.48 cm.		
SHOOTS AFTER EXCLUDING "ONE-CYCLE" SHOOTS			
A	37 cm.	37 cm.	31.5 cm.
K	0.471 ± 0.011	0.418 ± 0.020	0.361 ± 0.021
k	0.0127	0.0113	0.0115
Root-mean-square deviation of calculated values of x	0.55 cm.	0.74 cm.	0.67 cm.

If the group of meristematic cells in the apical bud of a lemon shoot are killed, growth is subsequently continued from a lateral bud in close proximity to the apex. Apparently the autocatalyst is diffused through many cells of the shoot, since when changed conditions bring lateral meristem into activity, it produces the same type of tissues as those produced by the apical meristem. So far as we can see, the physiological behavior of the shoots gives credence to the idea that the successive cycles of growth are due to the rhythmic activity of one and the same catalyst, having slightly smaller values in succeeding cycles and producing homologous structures in each.

The increase in size of the shoot was produced by the activity of a group of meristematic cells situated at the apex. These cells produced other cells which differentiated into various tissues, e.g., wood, leaves and buds, through which most of the materials necessary for growth were transported to the meristem. It is, therefore, apparent that the growth process of the meristem is regulated to some extent by one of the products of the reaction.

When the shoots begin their first cycle of growth in the spring, they draw largely on the substrates present in the tree, but as the cycle proceeds they supplement their supply of substrates with material drawn from the environment. The shoots may accumulate materials (especially during the latter part of the cycle) which serve as the substrate for the following cycle of growth. The amount of substrate accumulated depends to a large extent upon external factors, but the amount transformed by the plant seems to be more dependent upon internal factors. The processes of growth in succeeding cycles would be a repetition, to some extent, of the steps indicated for the first.

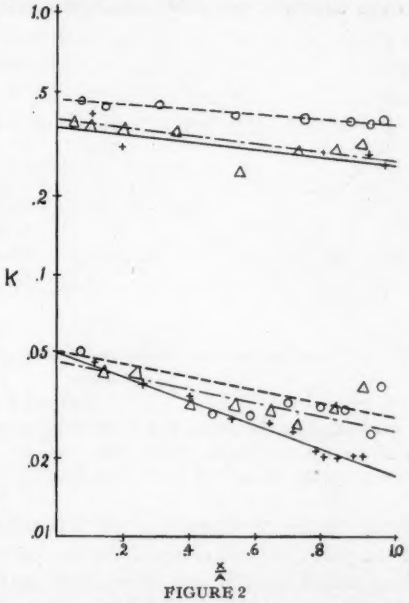


FIGURE 2

Values of K in successive cycles, showing the relation of log K to values of a/A in successive cycles. Uppes graphs, lemon shoots. Lower graphs, apricot shoots; data previously published.[6] ——— first cycle, — — — second cycle, ——— third cycle. Observed values; O O O first cycle, △ △ △ second cycle, + + + third cycle.

In the present state of our knowledge the analysis of the growth curves and their algebraic forms will supplement the analysis of the physiological data. The growth of these shoots shows a periodicity which is in harmony with the idea that the catalysis involved is a unimolecular reaction consisting of two consecutive reactions, one of which periodically retards the other. An attempt to describe the dynamical aspects of these consecutive reactions in plant growth has been made previously,[4,6] in which strong evidence of this very process was adduced. I examined the equation for growth proposed by Crozier,[2] which assumes that there may be two constants of the growth process, and calculated the values

of k_1 and k_2 for it. The values of the constants for the three cycles are given in table 3. The difference between the first and second cycles seems to lie chiefly in the values of k_1, the second and third differ in the values of A and of k_2. The interpretation waits for more detailed studies on the effects of environmental conditions. As yet there is no direct or

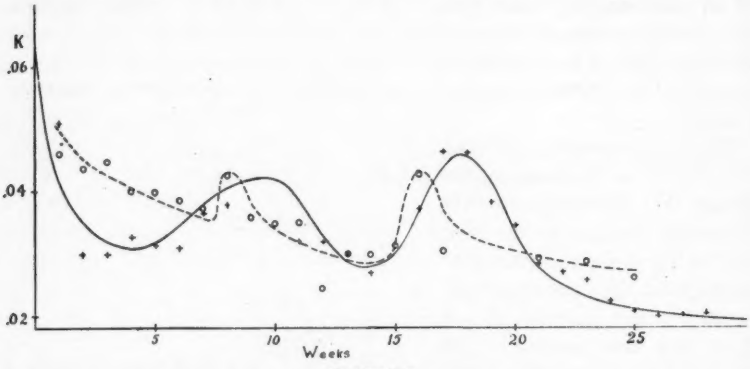


FIGURE 3

Oscillations in the values of K in three intra-seasonal growth cycles. ———, values of K for apricot shoots, taken from data in[6] — — — same x 10 for lemon shoots; + + + actual values of K for apricot shoots, ○ ○ ○ circles same for lemon shoots.

simple means of measuring the intensity of the retarding factor. One of the principal difficulties lies in the fact that many environmental conditions retard growth and their effect may mask that of the internal inhibitory agent. The nutrition of animals, however, may be controlled very accurately, and the occurrence of growth cycles in animals has frequently been reported.[1,10]

TABLE 3

CONSTANTS OF THE CURVES OF GROWTH OF THE LEMON SHOOTS FOR THE GENERAL EQUATION $t = \frac{1}{k_1 + k_2 A} \log \left(\frac{A(k_2 x + k_1)}{k_1 (A - x)} \right)$

	CYCLE I	CYCLE II	CYCLE III
A	36	35	26
k_1	0.014	0.0085	0.009
k_2	0.010	0.010	0.012
Root-mean-square deviation of calculated values of t (weeks)	0.086	0.398	0.698

The growth process of apricot shoots[6] likewise had three intra-seasonal cycles which could be expressed by the general equation now discussed. The work cited shows that the mean values of K also showed a progressive decrease in successive cycles, though the individual values rose to approximately similar levels at the commencement of the later cycles (Fig. 3).

The logarithmic values of K plotted against values of $\frac{x}{A}$ give approximately parallel straight lines (Fig. 2) and indicate that the factors which retard the growth process are approximately the same at a corresponding place in a cycle. It now seems that the ideas advanced in the paper just cited were only partially adequate to describe the activity of the catalyst, although it seems highly probable that the fluctuations in growth are due primarily to variations in the activity of the catalyst.

One-Cycle Shoots.—There were 17 shoots in the population under discussion which grew rapidly at first but ceased early in the season. They will be designated as "one-cycle" shoots. These one-cycle shoots were so well distributed throughout the population that they may be regarded as a normal phase of development of these trees. Their measurements were included in the averages thus far discussed and diminished somewhat the averages for the second and third cycles of growth of the entire lot.

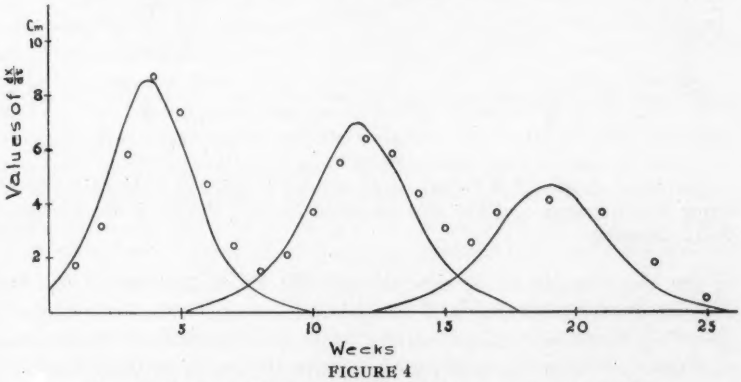


FIGURE 4

Increments of growth of lemon shoots plotted as values of dx/dt.

These one-cycle shoots reached a mean length of 28.11 cm. on the seventh week when the general population had a mean length of 33.72 cm. but the maximum length of the former was 28.93 cm. for the season, whereas the latter attained a length of 92.32 cm. Their growth is of interest because, although they were so closely parallel to the general population in the first four weeks of the season, they were completely arrested at the end of that time. When the values of K were plotted against $\frac{x}{A}$ the values of K for the one-cycle shoots were larger at corresponding stages of development than corresponding values of K for the general population (Fig. 5). This agrees with the fact that these one-cycle shoots completed their growth earlier than the remainder of the

population. The autocatalyst of growth in the three-cycle shoots resumed its activity in a short time and started a new cycle of growth. The constants of the equations for growth led one to the conclusion that the growth of the one-cycle shoots differed from that of their neighbors during the first cycle chiefly in the possession of a greater amount of autocatalyst. Some agency, however, terminated the activity of the autocatalyst at the end of the first cycle. Whether this agency is one of the products of the growth process, or whether it migrates into the shoots from other parts

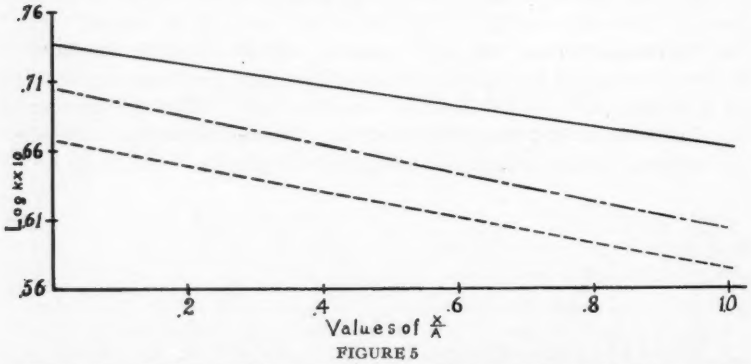


FIGURE 5

Comparison of values of K for one-cycle lemon shoots and those making three cycles. Upper, one-cycle shoots; middle, after excluding one-cycle shoots; lower, average of entire population.

of the tree remains to be determined. We know that starch is often associated with dormancy and its hydrolysis occurs when buds begin to grow. If conditions which favor the activity of carbohydrate-hydrolyzing enzymes also favor the activity of the growth catalyzers there might be renewed activity, resulting in the growth of a new cycle.

Summary.—There were three distinct cycles of growth of lemon shoots during a single season. During the first and second cycles the shoots made nearly equivalent amounts of growth, but they produced less in the third. The successive cycles of growth are assumed to be due to the periodic activity of a specific growth-promoting substance which catalyzed the growth process.

[*] Paper No. **182,** University of California, Graduate School of Tropical Agriculture and Citrus Experiment Station, Riverside, California.

[1] Brody, S., and A. C. Ragsdale, *J. Gen. Physiol.*, **3,** 623–633, 1921.

[2] Crozier, W. J., *Ibid.*, **10,** 53–73, 1926–27.

[3] Hedges, E. S., and J. E. Myers, *The Problem of Physicochemical Periodicity,* London, **1926,** pp. 95.

[4] Hooker, H. D., *Proc. Nat. Acad. Sci.*, **11,** 710–713, 1925.

[5] Reed, H. S., *J. Gen. Physiol.*, **2,** 545–561, 1919.

[6] Reed, H. S., *Proc. Nat. Acad. Sci.*, **6**, 397–410, 1920.
[7] Reed, H. S., *J. Agric. Res.*, **21**, 81–98, 1921.
[8] Reed, H. S., *Proc. Nat. Acad. Sci.*, **9**, 65–67, 1923.
[9] Reed, H. S., and Halma, F. F., *J. Agric. Res.*, **32**, 1177–1194, 1926
[10] Robertson, T. B., *J. Gen. Physiol.*, **8**, 463–507, 1926.

STUDIES ON THE BIOCHEMICAL DIFFERENCES BETWEEN SEXES IN MUCORS

4. Enzymes Which Act upon Carbohydrates and Their Derivatives

BY SOPHIA SATINA AND A. F. BLAKESLEE

CARNEGIE INSTITUTION OF WASHINGTON, COLD SPRING HARBOR, N. Y.

Communicated February 10, 1928

Our studies on the biochemical differences between sexes in Mucors have been carried along various lines during the past few years.[1,2,3] Most of the comparative tests which were previously made in an endeavor to discover the fundamental biochemical differences between the (+) and (−) races were quantitative in nature. In another series of tests, attempts have been also made to find out whether or not the (+) and (−) Mucor races could be distinguished qualitatively, whether, in other words, some substances which are present in races of one sex are absent in the races of the other sex. Some of the results obtained from such qualitative tests are presented in this paper.

A number of investigators have stated that the Mucor species which they tested are dimorphic in respect to certain qualitative biochemical differences. Thus, according to Breslauer[4] (Mucor hiemalis[5]), as well as Kostytschew and Eliasberg[6] (M. racemosus), only races of one sex contain sucrase while this enzyme is absent in the races of the other sex. This conclusion was based in each case on tests with one pair of races only. No evidence, however, is given to show whether the tested (+) and (−) races of M. hiemalis correspond to the same sexes of M. racemosus. In consequence, the results obtained by these authors are not directly comparable. It seemed desirable, therefore, to repeat the tests with a larger number of races not only to confirm the results just mentioned regarding the presence of sucrase in one sex and not in the other but also to discover how far such a qualitative biochemical difference could be demonstrated throughout this group of fungi. Since such a sexual difference might also be expected in respect to various carbohydrases, we made, in addition to tests for sucrase, a comparative study with a number of other carbohydrases, such as trehalase, amylase, emulsin and others which act upon

various carbohydrates and their derivatives. Twenty-four races, half of them (+) and half (−) were used in the tests and represent the following 10 species, included in 8 genera: Mucor mucedo, M. griseo-cyanus, Mucor V, Rhizopus nigricans (2 pairs), Absidia blakesleeana, Phycomyces blakesleeanus, Parasitella simplex, Circinella umbellata, Syncephalastrum racemosum and Cunninghamella bertholletiæ (2 pairs). These species were selected from various Mucor families and differ markedly from each other in physiological behavior. Thus they include a parasitic form (Parasitella), an extremely rapid grower (Cunninghamella) and other physiological types. The carbohydrases which were the object of this research are sucrase, trehalase, maltase, lactase, amylase, glycogenase, inulase and emulsin, the latter being considered as a mixture of enzymes acting upon various glucosides. The presence of the enzymes was tested with corresponding carbohydrates which are arranged in groups and listed on table 1. Tests with mannit were also made (see later).

TABLE 1

CARBOHYDRATES AND THEIR DERIVATIVES

MONOSACCHAR.	DISACCHAR.	POLYSACCHAR.	GLUCOSIDES	CARBOHYDRATE ALCOHOL
Glucose, 1% (used as control test)	Sucrose, 0.5–1%	Starch, 1%	Salicin, 0.5–1%	Mannit, 1%
	Trehalose, 0.5–1%	Glycogen, 0.5%	Amygdalin, 0.1–1%	
	Maltose, 1%	Inulin, 0.3–0.5%		
	Lactose, 1%			
		ENZYMES		
	Sucrase	Amylase	Emulsin	
	Trehalase	Glycogenase		
	Maltase	Inulase		
	Lactase			

The investigation was carried out with living cultures by inoculating the nutrient medium with a comparable amount of heavy spore suspension in sterile distilled water. The spores had previously been thoroughly washed in distilled water to remove traces of carbohydrates usually present in the unwashed spore suspension. Extreme care was taken to avoid infection and thus to prevent the possibility of an infecting microörganism converting the given carbohydrate. The cultures were incubated for 1–5 days at 26°C. The fungi were grown on a mineral nutrient solution (Czapec solution modified according to Harter and Weimer[7]) to which after inoculation the carbohydrate in question was added. In each case the given carbohydrate was the only source of carbon in the nutrient solution. To avoid a possible slight hydrolysis, during sterilization under

pressure of such carbohydrates as sucrose, salicin and inulin, they were sterilized by the discontinuous method at 100°C. For each experiment, 2 dishes were held as control. The first control contained the mineral nutrient solution, the second control contained the same solution to which an equal amount of a 1 per cent solution of glucose was added. Both dishes were inoculated with spores of the given species.

The spores in the first control dishes germinated well but never grew further, while in the second control dishes they developed a mycelium which, after a short period of time, formed a felt with aerial hyphæ and fructifications. The cultures in the experimental dishes were similar in appearance to those in the first control dishes in cases in which the fungus did not contain the hydrolytic enzymes capable of converting the carbohydrate which was added. Races, in which a suitable enzyme was present which hydrolyzed the carbohydrate, produced cultures which grew well and were similar in appearance to those growing in our second control dishes in which glucose was present as a source of carbon. These cultures when examined were graded with letters *A*, *B*, *C* according to the stage of their development. *A* signified a culture with aerial hyphæ and fructifications; *B*, a mycelium which had formed a well-developed felt or skin but no aerial hyphæ; *C*, hyphæ which had formed only separated flecks.

The presence of enzymes, therefore, could be determined directly by examination of the cultures.

In addition, the changes in the pH of the medium helped to demonstrate the presence of a carbohydrase since the acidity of the solution increases strikingly as the organism develops. Of diagnostic value also were certain other tests such as that with iodine which demonstrated the disappearance of starch and those with reagents which detect the presence of by-products like saligenin and hydrocyanic acid which accompany the formation of glucose. Thus, even when no reducing sugars could be detected in the medium, the action of emulsin could still be easily demonstrated, since saligenin gives a dark color reaction with $FeCl_3$, the intensity of color increasing with the growth of the fungus.

The presence of the reducing sugars in the nutrient solution was also tested in each case with either Fehling solution or by the MacLean volumetric method. When the enzyme was present and the growth of the fungus was slow, such a test was generally helpful in detecting the enzyme. In about 50 per cent of the cases, however, no traces of sugar could be detected in the nutrient solution. This was true, for the most part, in dishes which showed the best growth. Evidently the fungus used up the sugar as fast as it was formed. Purievisch[8] discovered a similar phenomenon when working with Aspergillus. In some cases sugar could be detected after the first day of growth, but disappeared later. In other

cases, however, it was only in well-developed cultures, several days old, that the first traces of sugar were found in the nutrient medium. Consequently, no definite conclusion regarding the occurrence of carbohydrases in the fungus can be drawn by testing for the presence of reducing sugars in the nutrient solution unless tests are made on successive days.

TABLE 2

CUNNINGHAMELLA BERTHOLLETIAE (+218, −216)

		SEX	24 HOURS	pH	48 HOURS	pH	72 HOURS
Trehalose pH 6.6	Growth stage	(+)	*B*−		*B*+		*A*−
		(−)	*B*		*A*−		*A*
	Reducing sugars in grams	(+)	0.015	3.7	0	3.6	0
		(−)	0.012	3.5	0	3.5	0
Sucrose pH 6.6	Growth stage	(+)	*C*		*B*+		*B*+
		(−)	*C*		*B*+		*B*+
	Reducing sugars	(+)	0.022	4.6	0.018	4.2	0.012
		(−)	0.024	4.6	0.020	4.0	0.016
Salicin pH 6.8	Growth stage	(+)	*C*+		*A*−		*A*−
		(−)	*C*+		*A*−		*A*
	Reducing sugars	(+)	0	4.4	0	−	0.006
		(−)	0	4.4	0	−	traces
	Saligenin	(+)	+		+		+
		(−)	+		+		+

Table 2 presents results of such tests with one pair of (+) and (−) races of Cunninghamella bertholletiæ which was grown on trehalose, sucrose and salicin. This is a species which grows extremely rapidly and develops its fructifications early. Cultures 24 hours old show a *B* stage of growth on trehalose, a *C* on sucrose and a *C*+ on salicin. The trehalose is evidently more easily assimilated than the other two substances. The changes of the pH values in the solutions confirm this conclusion since the lowest pH was in the medium with trehalose in which the speed of development was the most rapid. The reducing sugars are present in nutrient solutions to which trehalose or sucrose had been added but there are no reducing sugars in dishes with salicin. However, a test for saligenin with $FeCl_3$ shows clearly the presence of the emulsin. It is evident that a hydrolytic process has taken place in this case but that all the reducing sugar which was formed was utilized by the growing organism. Records taken from old cultures 48 and 72 hours old, show changes accompanying further development. The reducing sugar in media with trehalose disappears, that in media with sucrose decreases, while, on the other hand, reducing sugar makes its appearance in solutions with salicin only after the development of the fungus has been completed.

In table 3 is given a summary of the results obtained by the same methods from all the 24 (+) and (−) races mentioned above which were tested with the 10 carbohydrates and their derivatives listed in table 1. The data concerning the grade of growth, the amount of reducing sugar present, the pH value and the tests with the accompanying by-products have been confirmatory of the conclusions reached from the development of the fungus but are details which need not be presented here. The + sign on table 3 indicates presence of the enzyme which converted the corresponding carbohydrate added to the nutrient solution, the 0 sign the absence of such an enzyme. As it appears, the trehalase, maltase, emulsin, amylase and glycogenase are widely distributed in the Mucors. Less common is the lactase while sucrase and inulase were found only in two species. No one of these enzymes was limited to one sex in any of the species tested. If the (+) race of a species was able to convert the carbohydrate, the same was true of the (−) race. If the enzyme was absent in the race of one sex of the species tested, it was absent in the race of the other sex. The sample of species tested shows that it is not a general

TABLE 3

	SEX	GLUCOSE	TREHALOSE	MALTOSE	SALICIN	AMYGDALIN	STARCH	GLYCOGEN	LACTOSE	SUCROSE	INULIN	MANNIT
Cunninghamella 217, 266	(+)	+	+	+	+	+	+	+	0	+	+	+
	(−)	+	+	+	+	+	+	+	0	+	+	+
Cunninghamella 218, 216	(+)	+	+	+	+	+	+	+	0	+	+	+
	(−)	+	+	+	+	+	+	+	0	+	+	+
Absidia blakesleeana	(+)	+	+	+	+	+	+	+	+	0	0	+
	(−)	+	+	+	+	+	+	+	+	0	0	+
Circinella umbellata	(+)	+	+	+	+	+	+	+	+	0	0	+
	(−)	+	+	+	+	+	+	+	+	0	0	+
Syncephalastrum	(+)	+	+	+	+	+	+	+	+	0	0	+
	(−)	+	+	+	+	+	+	+	+	0	0	+
Rhizopus 43, 309	(+)	+	+	+	+	+	+	+	+	0	0	+
	(−)	+	+	+	+	+	+	+	+	0	0	+
Rhizopus 302, 303	(+)	+	+	+	+	+	+	+	+	0	0	+
	(−)	+	+	+	+	+	+	+	+	0	0	+
Phycomyces blakesleeanus	(+)	+	+	+	+	+	+	+	0	+	0	0
	(−)	+	+	+	+	+	+	+	0	+	0	0
Mucor 5	(+)	+	+	+	+	+	+	+	0	0	+	0
	(−)	+	+	+	+	+	+	+	0	0	+	0
Mucor mucedo	(+)	+	+	+	+	+	+	+	0	0	0	+
	(−)	+	+	+	+	+	+	+	0	0	0	+
Mucor griseo-cyanus	(+)	+	+	+	+	+	+	+	0	0	0	0
	(−)	+	+	+	+	+	+	+	0	0	0	0
Parasitella simplex	(+)	+	+	+	+	+	0	0	0	0	0	0
	(−)	+	+	+	+	+	0	0	0	0	0	0
Total		24	24	24	24	24	24	22	22	10	6	6	16

rule in Mucors that one sex has an enzyme which is lacking in the opposite sex. It is difficult to understand the results of Breslauer and of Kostytschew, unless they are due to their having worked with only a single pair of races. It is possible that tests of a larger number of races might disclose individuals which would react differently from others of the same sex, as we have found to be the case in earlier investigations.[1,2,3]

Our experiments permit us to draw some conclusions regarding the relative value of different carbohydrates for Mucor growth. These are arranged in table 3 according to their relative value. Amygdalin should be excluded from this list since this cyanophoric glucoside yields hydrocyanic acid as one of the by-products of hydrolysis. All the cultures with amygdalin, in consequence, started to develop very well and produced a large amount of reducing sugars but were soon killed by this toxic substance after a few days of growth. The trehalose, maltose and salicin are very favorable for the fungi and can be substituted in cultures for the glucose. All of the 24 tested races grew well on nutrient medium containing these substances and in some cases the growth of the fungi was even more vigorous than in the second control dishes with glucose. The same was true of the starch and glycogen. The only species which did not convert them was the parasitic form Parasitella.

Much less valuable are lactose, sucrose and inulin which were assimilated by but few species. The growth of Rhizopus on lactose was very feeble but since it failed to grow in the first control, a slight utilization of this carbohydrate is indicated. Without going into details, it may be mentioned that others have found a similar situation among the Mucors in respect to the relative values of the different carbohydrates.

Besides tests with the other carbohydrates listed in tables 1 and 3, the 24 (+) and (−) races were also grown on media to which the carbohydrate alcohol mannit was added. According to Czapec,[9] nothing definite is known concerning the utilization of the mannit. For this reason, the results with mannit are given separately. Mannit, trehalose and glycogen are the most common and important carbohydrates in fungi. In some cases the amount of mannit in higher fungi reaches as much as 20 per cent.[9] The reports, however, concerning the value of mannit as a general source of carbon are very contradictory. In certain groups of fungi, mannit has been reported by some investigators as being assimilated but by other investigators as not available.[9] Our tests which were applied to races of several species within a single group also show that some of the species can and some cannot assimilate mannit. Thus 6 species out of 10 were able to utilize mannit successfully (Cunninghamella, Absidia, Circinella, Syncephalastrum, Rhizopus and Mucor Mucedo). The other 4 species (Mucor V, M. griseo-cyanus, Phycomyces, Parasitella) did not develop on medium in which mannit was the only source of carbon. In all, 16

races out of 24 could grow on this substance. In all cases, however, there was no sexual difference in respect to utilization of mannit.

The Mucor species in table 3 are arranged according to their behavior with the various carbohydrates and their derivatives. Cunninghamella is able to convert all the carbohydrates tested except one (lactose). On the other hand, the three species of the genus Mucor were less adaptable as regards the carbohydrate upon which they could grow. Out of 10 carbohydrates and their derivatives tested, 3 (4 in case of M. griseocyanus) were not available as sources of carbon. As might have been expected, the parasitic form Parasitella was the most limited in respect to the kinds of carbohydrates upon which it could grow, being able to assimilate only four of the sources of carbon which were tested.

Summary.—1. Comparative tests were made with 24 (+) and (−) Mucor races (included in 10 species and 8 genera) to determine the possibility of a given carbohydrase being present in one sex and not in the other.

2. No such qualitative differences between the sexes were discovered.

3. Trehalase, maltase and emulsin were present in all the tested species; amylase and glycogenase in all except one. Lactose was found in races of 4 species and sucrase and inulase in races of 2 species only.

4. Among the species tested, Cunninghamella appeared to contain the greatest numbers of carbohydrases, species of the genus Mucor and Parasitella contained the fewest.

5. In addition to glucose used as control, the relative value for Mucor development of 10 carbohydrates and their derivatives, including di- and poly-saccharides, glucosides and a carbohydrate alcohol, can be shown approximately by the following order: Trehalose, maltose, salicin, amygdalin, starch, glycogen, mannit, lactose, sucrose and inulin. Amygdalin, however, yields hydrocyanic acid and ultimately kills the fungus.

[1] Satina, S., and Blakeslee, A. F., these PROCEEDINGS, **11**, 528–534 (1925).

[2] *Ibid.*, **12**, 191–202 (1926).

[3] *Ibid.*, **13**, 115–122 (1927).

[4] Breslauer, A., *Bull. Soc. Bot. Genève*, 2 ser., **4**, 228–237 (1912).

[5] Lendner, A., *Ibid.*, 2 ser., **11**, 70–84 (1918). Lendner, however, presents evidence that both the (+) and (−) races of this same species are able to invert sucrose.

[6] Kostytschew, S., and Eliasberg, P., *J. Russ. Bot. Soc.*, **4**, 39–40 (1919).

[7] Harter, L., and Weimer, J., *J. Agric. Res.*, **20**, 761–786 (1921).

[8] Purievisch, K., *Ber. Beut. Bot. Gesel.*, **16**, 368–377 (1898).

[9] Czapec, F., *Biochemie der Pflanzen*, Vol. **1**, 308–309 (1922).

INDUCED IMMUNITY IN PLANTS

BY DONTCHO KOSTOFF

BUSSEY INSTITUTION OF HARVARD UNIVERSITY

Communicated February 15, 1928

A series of experiments on the family Solanaceæ has apparently shown that plant tissues can acquire immunity against certain antigenic agents of other species by some type of antibody production.

Normal precipitins were found in leaf extracts and stem extracts and these were tested by Uhlenhuth's ring method (1909).

A great variability in amount of precipitation was observed when normal plant extracts were tested against each other. Powerful precipitation was observed both when the two extracts tested against each other were from plants of the same genus and when plants of different genera were used; but in certain combinations—particularly with different genera—instead of a precipitation ring, a lytic ring (a ring of liquid clearer than the two components) was observed.

The precipitation potency of certain species and genera is increased after grafting; moreover, in certain combinations whose extracts show no precipitin reaction before grafting, the capacity to produce precipitins is acquired during the growth of the graft unions. The highest acquirement of precipitins was found between 30 and 45 days after grafting. Acquirement of precipitins after grafting was observed in the following graft unions: *Nicotiana Rusbyi* grafted on *Nicotiana rustica*, *Solanum tuberosum* grafted on *Nicotiana rustica*, *Nicotiana Sanderæ* grafted on *Datura ferox*, *Nicotiana glauca* grafted on *Capsicum pyramidale*, *Lycium Barbarum* grafted on *Solanum Lycopersicum*, etc.

The precipitins acquired after grafting are usually specific. Thus extract of *Nicotiana Rusbyi* immunized against *Nicotiana rustica* by grafting always gave the heaviest precipitate when tested against its antigenic extract, namely, the normal extract of *Nicotiana rustica*. Then followed in order the extracts of *Nicotiana paniculata*, *Nicotiana glauca*, F_1 (*Nicotiana rustica* × *Nicotiana Tabacum*) and *Nicotiana Tabacum*. It is perhaps important to note in this connection that the three species *Nicotiana rustica*, *Nicotiana paniculata* and *Nicotiana glauca* belong to the subgenus or section *Rustica*, while the species *Nicotiana Rusbyi* and *Nicotiana Tabacum* belong to the section *Tabacum* (Comes, 1899).

Acquired precipitins of *Solanum tuberosum* against (grafted on) *Nicotiana rustica* precipitate the normal extract of *Nicotiana rustica*, but not the normal extract of *Nicotiana Rusbyi*. Sometimes, however, the extracts of *Nicotiana Rusbyi* immunized by grafting against *Nicotiana rustica* showed increased precipitation potency over the precipitation shown

against the normal extract of *Nicotiana alata*. The latter *Nicotiana* species belongs to the section *Petunioides* and the acquired precipitins of *Nicotiana Rusbyi* cannot be called "*specific*" but "*heterogenetic.*"

In certain graft unions no acquirement of precipitins was found;, viz., *Nicotiana Tabacum* grafted on *Nicotiana rustica*, *Nicotiana rustica* grafted on *Nicotiana quadrivalvis*, *Nicotiana Tabacum* grafted on *Solanum nigrum*, *Salpiglossis sinuata* grafted on *Datura ferox*, *Solanum Melongena* grafted on *Solanum nigrum*, *Nicotiana Tabacum* grafted on *Solanum Melongena*, *Nicotiana Tabacum* grafted on *Solanum tuberosum*, *Nicotiana glauca* grafted on *Nicotiana rustica*, etc.

A decrease of the normal precipitin potency was observed in the following graft unions: *Solanum tuberosum* grafted on *Datura Wrightii*, *Solanum Lycopersicum* grafted on *Datura Wrightii*, *Solanum Melongena* grafted on *Datura Wrightii*, etc.

Comes, O., Monographie du genre Nicotiana, Topographie coöperative, Naples, 1899.

Uhlenhuth, P. u. Weidanz, Technik und Methodik des biologischen Eiweissdifferenzierungsverfahren u.s.w., Gustav Fischer, Jena, 1909.

ON THE BINOCULAR FUSION OF COLORS AND ITS RELATION TO THEORIES OF COLOR VISION

BY SELIG HECHT

LABORATORY OF BIOPHYSICS, COLUMBIA UNIVERSITY

Communicated January 24, 1928

1. Thomas Young's original idea for the mechanism of color vision rests on the fact that by mixing three selected monochromatic parts of the spectrum all known color sensations may be reproduced. Young[1] supposed that there are three kinds of "fibers" in the retina each producing a characteristic sensation, one of red, another of green and a third of blue (violet). Each type of fiber is sensitive practically to the entire visible spectrum, but the first possesses a maximum of sensibility in the red, the second in the green and the third in the blue. The various color sensations then result from the relative strengths with which the three different fibers are stimulated by the objective light.

Few people today suppose that Young's idea, as here given in its simple form or as elaborated by Helmholtz,[2] Koenig,[3] v. Kries[4] and others, is adequate as a complete theory for the mechanism of color vision. The question, however, has often been raised as to whether it can serve even as the basis for a theory of color vision.

The main objection to Young's hypothesis is founded on the uniqueness of the sensations of yellow and of white. Mixtures of green and blue lights give a continuous series of colors in which both are identifiable. The same is true with mixtures of red and blue lights. But in mixtures of red and green a new sensation, yellow, arises which contains neither red nor green. The same is true for mixtures of yellow and blue, or of green, red and blue, which give white, a sensation *sui generis*. If Young's idea is correct then it must be supposed that yellow is a central phenomenon which occurs when in the retina red and green receiving fibers function simultaneously. Similarly, white occurs in the brain when all three—red, green and blue—receiving fibers function in the retina. It is precisely against such a formulation that criticism has been raised by Hering[5] and more recently by Ladd-Franklin.[6] And as result, first Hering and later Ladd-Franklin have devised theories which assume separate receptor substances for yellow and for white.

If there is to be developed an adequate theory for the mechanism of color vision a decision must be made in the very beginning as to which of these two conceptions of yellow and white is correct. Are there special substances in the retina for the reception of yellow and white, or are yellow and white phenomena which arise in the brain out of the impulses coming from the three kinds of fibers (or substances) in the retina? The answers to these questions must be given experimentally.

If red light and green light fall on a given retinal area of one eye and a yellow sensation results, it is not possible to decide among the following interpretations of what happens: (*a*) The photochemical product of the action of red light combines chemically with that formed by green light to give a "yellow" substance which stimulates the proper nerve endings, as Ladd-Franklin postulates; (*b*) a substance reversibly sensitive to red and green remains unchanged while a substance reversibly sensitive to yellow and blue is changed by the red light to which it is partially sensitive, as Hering proposed; (*c*) the red sensitive fibers and the green sensitive fibers are both active and the brain synthesizes yellow, as must be the case if the Young-Helmholtz idea is adopted. But if red light falls on a part of the retina of one eye, and green light falls on the corresponding portion of the retina of the other eye and the result is a yellow sensation, then (*a*) and (*b*) cannot be true, and only Young's idea is tenable. This is precisely what happens in the binocular mixing of colors.

2. The problem of the binocular mixing of colors is an old one; it had been a subject of controversy before Helmholtz. But it was Helmholtz in his "Physiologische Optik" who settled the matter temporarily by stating that he was unable to fuse colors binocularly. This has been generally accepted even though Hering[7] later showed the cause of Helmholtz's failure and arranged an apparatus to overcome it. As a result of

the further suggestions of Schenck[8] and of Stirling,[9] one finds v. Kries in the 1911 edition of Helmholtz's "Physiologische Optik" admitting the possibility of binocular color fusion, though he finds it difficult to accomplish.

Since then Trendelenburg[10] has constructed a binocular modification of the classical Helmholtz monocular color-mixing apparatus and has demonstrated beautifully that not only can binocular color fusion take place, but that one can obtain regular color mixing equations comparable to those obtained monocularly. Red and green binocularly give yellow; and several complementary pairs give white. However, even with this improved apparatus the process is described as difficult and requiring considerable practice. Recently Rochat[11] has devised another type of spectroscopic apparatus for binocular color mixing and obtained the same results as Trendelenburg. Here again emphasis is laid on the difficulties and practice involved. Dawson[12] has also reported binocular mixtures, but with none of the elegance of Trendelenburg and Rochat.

It must, therefore, be concluded that modern work has definitely settled the problem of binocular color mixing. Red in one eye and green in the other on identical retinal areas result in a yellow sensation. This yellow obviously cannot arise anywhere but in the brain. The only difficulty seems to be that binocular color mixing is reputed to be a hard feat to perform and to require much practice. The purpose of this paper is to report that having evolved a device for testing the binocular fusion of colors unprejudiced as to its difficulties because of my ignorance of the previous work, I found it to be a simple experiment which requires no practice at all.

3. The easiest method of demonstrating binocular color mixing is to put a red filter in front of one eye and a green filter in front of the other and to look at a brightly illuminated white surface about 20 cm. square placed on a black background. It is essential that the two filters transmit light of about the same brightness. Wratten Filters 29 (red) and 58 (green) are excellent. With one eye open the surface appears red; with the other open it appears green; with both open it is yellow.

This is the simplest but not the most effective method. It is better to see the three colors side by side—red, green, and between them, the centrally mixed yellow—and to compare them with one another. A blackened box, 30 cm. long, 15 cm. wide and 10 cm. high is arranged so that one 15 × 10 end is open and fits against the face. The other 15 × 10 end is closed, and contains the two 5 cm. square Wratten Filters, one next to the other, and in a light-tight fit. About a meter from one's eyes is a white cardboard with a bright light as near in front of it as possible. We use a 250-watt or a 100-watt concentrated-filament lamp projecting through a hole in the cardboard. This illuminates the cardboard and serves as a point of almost forced fixation for the two eyes. On looking

at the bright light with *both* eyes focussed on it through the box, one immediately sees a bright yellow light in the center of a yellow square of cardboard, flanked on one side by a green square and on the other by a red square.

What actually happens one can determine by looking with one eye at a time. One eye looks at the light through the green filter, and at a neighboring piece of cardboard through the red filter. The other eye looks at the light through the red filter and at a neighboring piece of cardboard on the other side through the green filter. By focussing both eyes on the light the red image in one eye and the green image in the other eye fall on identical points of the two retinas and are fused in the brain to give the appearance of a yellow light. The neighboring green in one eye and red in the other cannot be made to overlap at the same time and, therefore, are seen monocularly as green and red. The total sensation is somewhat as if one were looking at the light and the cardboard through three colored openings, a red (monocular), a yellow (binocular) and a green (monocular). The effect appears instantaneously if the eyes are focussed correctly.

No explanation, except the injunction that both eyes look at the light, is necessary for performing this binocular fusion. During last summer at Woods Hole, when the apparatus in a slightly different form was used, about a hundred people, trained and untrained, young and old, performed the experiment with no practice or difficulty. Six individuals failed to make the proper responses. Of these, five turned out to be color blind. The sixth left after a few moments' trial and, unfortunately, did not return, so that it could not be determined whether he is color blind or not. Since setting up the experiment in New York about twenty more people have tried it with uniformly successful results.

The green and red Wratten Filters in the box may be replaced by yellow and blue filters to make binocular white. Wratten filters 16 (yellow) and 44A (blue) have been the best combinations thus far, and give a reasonably good white.

The binocular formation of yellow and of white, therefore, shows that such theories as those of Hering and of Ladd-Franklin, which require special receptors for yellow and white, are neither necessary nor correct. The uniqueness of yellow and of white as sensations is no obstacle to adopting Thomas Young's three-receptor idea as a basis on which to build a theory for the mechanism of color vision.

[1] Young, *Lectures in Natural Philosophy*, ii, London, **1807.**

[2] v. Helmholtz, *Handbuch der Physiologischen Optik*, 2nd edition, Hamburg and Leipsic, **1896.**

[3] Koenig, *Gesammelte Abhandlungen zur physiologischen Optik*, Leipsic, **1903.**

[4] v. Kries, in Nagel's *Handbuch der Physiologie des Menschen*, iii, Braunschweig, **1905.**

[5] Hering, *Sitzungsber. k. Akad. Wiss., Vienna, Math.-Naturw. Cl.*, **69**, 1874.
[6] Ladd-Franklin, *Zeit. Psychologie*, **4**, 1893; *Science*, **55**, 1922.
[7] Hering, in Hermann's *Handbuch der Physiologie*, iii (1), Leipsic, **1879**.
[8] Schenck, *Sitzungsber. phys.-med. Gesellsch.*, Wurzburg, **1898**.
[9] Stirling, *J. Physiol.*, **27**, 1901.
[10] Trendelenburg, *Arch. ges. Physiol.*, **201**, 1923.
[11] Rochat, *Arch. néerland. Physiol.*, **7**, 1922.
[12] Dawson, *British J. Psychol.*, **8**, 1917.

ON THE MOTION OF THE MAGELLANIC CLOUDS

BY WILLEM J. LUYTEN

HARVARD COLLEGE OBSERVATORY

Communicated February 14, 1928

The Magellanic Clouds appear to be unique in that they are the only objects outside the Galactic System proper for which we may expect to determine the total space motion with respect to the galaxy within some reasonable time.

Thus far, all that we know definitely concerning them, is:

(*a*) The radial velocities of eighteen gaseous nebluæ in the large cloud, and of two in the small cloud.[1]

(*b*) The distance, based upon the behavior of certain variables in the clouds. The most recent values are those of Shapley,[2] who gives 3.2×10^4 parsecs for the small cloud, 3.4×10^4 for the large cloud, both affected by an error estimated to be not more than fourteen per cent.

The apparent proximity of the clouds in the sky, and their almost identical distance constitute prima facie evidence that the assumption that the two clouds are physically connected is not improbable. Hertzsprung started from this point and, assuming that a perspective effect was the cause of the progressive change in the radial velocities in the large cloud (from 305 km. sec.$^{-1}$ to 252 km. sec.$^{-1}$) and of the very different radial velocity of the small cloud (163 km. sec.$^{-1}$) calculated the total space motion of the two clouds. His values for the apex of this motion and its speed are

$$\text{R. A. } 4^h41^m \pm 19^m, \text{ Dec. } -5.6° \pm 3.0°, \text{ V } 620 \pm 70 \text{ km. sec.}^{-1}.[3]$$

Naturally, Hertzsprung has himself computed the angular proper motions to be expected on this theory. Using Shapley's later value for the distance of the clouds, and indicating by λ the angular distance of the clouds from their apex, by p the direction of their proper motion, and by T their tangential velocity relative to the sun, we obtain:

	λ	T(km.s)	T(A.U.)	p	$\mu\alpha \cos\delta$	$\mu\delta$
Small cloud	72	600 ± 65	126 ± 14	55	$+0.003''$	$+0.002$
Large cloud	64	560 ± 70	118 ± 15	348	$-0.0007''$	$+0.004$

Hertzsprung has already mentioned several ways of determining these proper motions, and the following methods should be considered merely as an extension of his suggestions.

1. The motion of stars known to belong to the clouds may be determined directly by either of two methods:

(*a*) Stars of the tenth magnitude or brighter may be directly observed with the meridian circle. Such stars could be selected as soon as radial velocities were available, or they might be found from objective prism spectrograms by reversing the prism, since the high velocity would produce a measurable effect, especially for stars in the large cloud. In view of Shapley's work it does not seem unreasonable to expect to find at least a dozen stars of this type.

(*b*) Proper motions may be determined for variable stars of the typical Magellanic Cloud type. The median magnitudes of the majority of these stars are between fifteen and seventeen. Their systematic motion must be determined relative to foreground stars for which the absolute motion is known. This necessitates the use of an objective grating by means of which the relative photographic proper motions of these faint variables may be connected, through three or four successive steps, with the motion of stars of the ninth magnitude or brighter, for which meridian proper motions have been determined. At present hardly any meridian proper motions are known in this region of the sky. Fortunately, the La Plata observatory has undertaken the observation of all stars in the C. P. D. brighter than the ninth photographic magnitude and between the declination limits $-57°$ and $-82°$, and according to the statement made in the introduction of the LaPlata C catalogue the observation of these stars is now practically completed. If they are re-observed fifty years hence, the resulting proper motions may be expected to have an individual mean error of about $0.01''$, and as it should be possible to use between fifty and one hundred stars of the foreground, the mean accidental error of the results should be of the order of $0.001''$ to $0.002''$, an accuracy sufficient only to indicate whether the predicted value for the motion of the clouds is of the right order of magnitude. If method (*a*) is used, fundamental observations must be made, and a much longer time must elapse before a mean value derived from a dozen stars can have the required accuracy.

2. The motion of the cloud stars may be determined relative to faint foreground stars for which the mean systematic drift motion relative to the sun may be considered as known.

(*a*) First among such foreground stars are faint proper motion stars

such as those between magnitudes fourteen and sixteen with motions between 0.02″ and 0.04″ annually. If objective-prism spectra of such stars are obtained, their estimated absolute magnitudes, in connection with the known velocity of the sun may help to evaluate the exact amount of their parallactic drift. Unfortunately, it appears, judging from preliminary evidence derived with the blink microscope for stars in front of the large cloud, that the motion of a number of these stars is not towards the usually accepted solar antapex.

(*b*) Proper motions may be measured for stars for which the mean motion relative to the sun is so small that in spite of their much smaller distance their mean angular motion is of a lower order of magnitude than the motion of the cloud. The only faint stars that can be considered are stars of spectral class *A* (*B* stars are statistically absent in the high galactic latitude of the clouds and *K* giants could not easily be distinguished from *K* dwarfs). The writer has previously shown[4] how class *A* stars may be used to determine the systematic errors of proper motions. For the Magellanic Clouds we should have to go beyond the twelfth magnitude before reaching stars of proper motion less than 0.002″ annually, and again the method becomes impracticable on account of the few stars available—a dozen or so at best.

It is an unfortunate coincidence that the probable apex of the Magellanic Clouds is only thirty degrees distant from the solar antapex, thus causing the parallactic drift of the foreground stars to be in nearly the same direction as the motion of the cloud.

Summing up we are forced to conclude that the prospect of actually determining the motion of the Magellanic Clouds is not very promising. Nevertheless, it seems to the present writer that all methods should be given a trial, since a combination of them should, in fifty years' time reveal to us at least the upper limit of the proper motion. It should be emphasized that this pessimistic forecast is wholly dependent upon the distance adopted. If Shapley's distance should prove to be much too large the actual proper motions would be correspondingly magnified, and if Hertzsprung's original value of 10^4 parsecs is correct, the resulting proper motions would be of the order of 0.012″ annually, and twenty-five years' observing might well reveal them.

The available knowledge of the Magellanic Clouds is meager, and even that little has now been exhausted. Yet it is tempting to continue and walk the path of pure extrapolation and speculation, especially concerning the relation of the Magellanic Clouds to the Galaxy. It is perhaps not unreasonable to assume that the total mass of the Galaxy is at least of the order of 5×10^{10} times that of the sun. It seems more than equally reasonable to suppose that the combined mass of the two clouds is considerably less than one-tenth of this value. Let us further suppose that

Hertzsprung's value for the total motion of the clouds is approximately correct, and not vitiated by a large K-term affecting the radial velocities or by a relative motion of the two clouds.* Let us represent the two clouds by one point in which all their mass is concentrated and situated one-third of the distance from the large cloud to the small one. The value of the radial velocity at this point is 238 km. sec.$^{-1}$ = 50 astr. units a year; the tangential velocity is 117 A.U. ann.$^{-1}$. Finally, let us assume that the motion of the sun with respect to the center of gravity of the galactic system is small and that as a very rough first approximation all the mass of the Galaxy may be regarded as concentrated in the sun.† Under these almost ideally simple conditions we may now calculate the orbit of the Magellanic Clouds around this fictitious center of the Galaxy. The orbit proves to be a hyperbola of high eccentricity; its constants are given below in the column headed (*a*). To show that this result is very sensitive to changes in the adopted constants of the system Galaxy–Clouds, similar calculations have been made under the assumptions (*b*) that the combined fictitious mass of the systems is ten times as large, and (*c*) that, in addition, the distance between us and the clouds is three times smaller. These suppositions, especially the second one, appear at present rather improbable, but they are by no means impossible. It will be noticed that, in case (*b*) the orbit, though still a hyperbola, is of much lower eccentricity, corresponding to an angle of 150 degrees between the asymptotes, and that in case (*c*) the orbit has become elliptical.

M	(Sun)	5×10^{10}	5×10^{11}	5×10^{11}
r	(Astr. units)	6×10^{9}	6×10^{9}	2×10^{9}
a	(Astr. units)	-1.3×10^{8}	-2×10^{9}	$+5 \times 10^{9}$
e		45	3.7	0.63
α		177.5°	149°	....
P	(years)	.....	.....	5×10^{8}

These speculations, already carried far beyond the limits of credulity usually endured, suggest the unwisdom of making any further statements concerning the place and time of "perigalacton." The computations are given merely to indicate that the path of the Magellanic Clouds with respect to the Galaxy may deviate appreciably from a straight line, the mutual influence of "island universes" upon each other's paths being considerably more pronounced than among individual stars in the galactic system. It thus appears within the realm of possibilities that the Magellanic Clouds, though somewhat vagrant and "cometary" in behavior, are permanent members of our Galaxy.

* Assuming the combined mass of the clouds to be 10^9, the parabolic velocity of the small cloud relative to the large is 25 km. sec.$^{-1}$, and the foregoing assumption appears to be slightly dangerous in the most favorable case.

† If the center of the galactic system assumed to be in longitude 325° and at a distance of 6000 parsecs is adopted as center of reference, and the velocity of the sun

assumed to be 286 km. sec.$^{-1}$ in direction of 55° longitude[5] is allowed for, very little change is brought about in the results of the computation. The velocity of the clouds then proves to be 90 km. sec.$^{-1}$ radially, and 583 km. sec.$^{-1}$ tangentially, with respect to this center and no appreciable change takes place in the constants of the hyperbolic or elliptic orbits. The plane of the orbit, however, would be altered, since this is in all cases supposed to pass through the velocity vector and the center of reference.

[1] *Lick Obs. Publ.*, **13**, 1923.

[2] *Harvard Circular*, No. **255**, 1924; No. **268**, 1924.

[3] *Mon. Not. R. A. S.*, **80**, 1920, 782; **83**, 1923, 348.

[4] *Harvard Bulletin*, No. **801**, 1924.

[5] Oort, *Bull. Astr. Inst. Neth.*, Nos. **120** and **132**, 1927.

ARTIFICIAL PRODUCTION OF JANUS EMBRYOS OF CHAETOPTERUS

BY ALBERT TITLEBAUM

DEPARTMENT OF ZOÖLOGY, COLUMBIA UNIVERSITY

Communicated January 19, 1928

The eggs of Chaetopterus were compressed between a slide and coverslip, at various stages ranging from prior to fertilization to the four-cell stage. It was found that when pressure is applied at a stage shortly after the extrusion of the second polar body (the "pear-shaped" stage), and released when the cleavage plane has cut about half-way through the egg, many of the eggs cleave equally. Normally the first cleavage is unequal, and the "polar lobe," characteristic of this egg, returns to the larger (*CD*) blastomere (Fig. 1*a*). But in the equally cleaving eggs, the material of the first "polar lobe" goes to both blastomeres (Fig. 1*b*). In some of these eggs the first cleavage plane disappears and, if no subsequent cleavage occurs, they sometimes later show "differentiation without cleavage." If other cleavages take place abnormal larvæ are formed which usually die in two days. However, those equally cleaving eggs, in which the first cleavage persists, continue to cleave and many develop later into double embryos, mainly of the duplicitas cruciata type (Fig. 2*b*).

The second cleavage of these eggs gives two larger and two slightly smaller cells, two very small "polar lobes" being sometimes seen on each of the two larger cells, instead of the single second "polar lobe" which normally goes to the *D*-blastomere. The third division seemed to be in the same direction (dexiotropic) as in the normal egg.

The early trochophores from these equally cleaving eggs appear to be double structures, but might be confused with other abnormal types. However, the later transparent trochophore shows its double nature unmistakably (Fig. 2*a*, *b*). The most complete ones have four eye spots and

a single apical flagellum. Two mouths lead into a single large stomach which connects with the intestine of each embryo. Each embryo has two lateral flagella, a second ciliary band about each intestine, and a "holdfast." The head-ends are fused back to back and the trunks are separate, the median plane of the fused heads being at right angles to the median plane of the two separate trunks, giving the Janus type. In most cases one embryo was usually complete, while the other ranged from completeness to various stages of imperfect development of the holdfast, intestine, ciliary band, two flagella, eye spots, etc.

Eggs, compressed for different lengths of time and at other stages up to the first cleavage occasionally give double embryos but in much fewer numbers. When pressure is applied from the two- to the four-cell stage the *CD* cell frequently divides equally, but the embryos usually do not

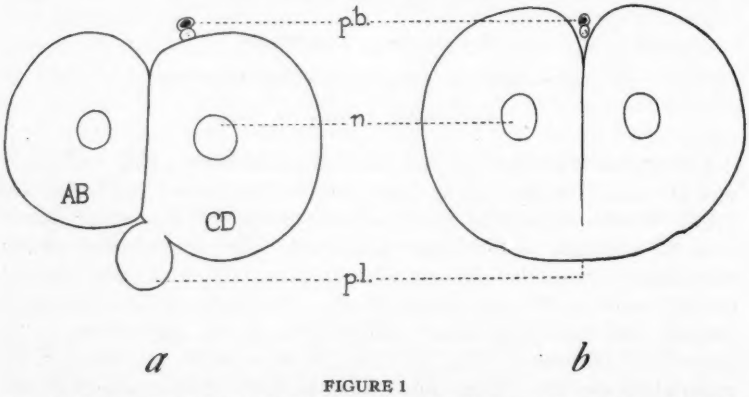


FIGURE 1

Fig. 1(*a*) Normal two cell stage; (*b*) equally cleaving egg; n, nucleus, p.b., polar bodies, p.l., polar lobe.

develop beyond the early trochophore stage, at which time double embryos are very difficult to distinguish. When pressure is applied at the four until the eight-cell stage at right angles to the polar axis, the *D* cell, as well as each of the others, divides equally, and though some of the larvae resemble the double forms they do not develop far enough to be identified with certainty. In all cases the cleavage planes were at right angles to the compressing coverslip.

The formation of the double embryos appears here to be correlated with the equal distribution of the substance of the "polar lobe" to the two blastomeres in the first cleavage. Penners ('24) interpreted his double embryos in Tubifex as due to the distribution of the "pole plasms" to each of the first two blastomeres. But for the present, the more probable interpretation is that perfect Janus embryos develop only in those cases

where the egg happens to be compressed in such a way that the first cleavage spindle develops at right angles to its normal position, though still at right angles to the polar axis of the egg. The resulting plane of cleavage then divides the entire egg, including the "polar lobe" material, into two equal parts.

The blastomeres of equally cleaving eggs were isolated with a glass needle, and in several instances whole embryos (late trochophores) were formed. Both blastomeres of the same egg were followed in many cases, but usually one or both died at the early trochophore stage where it is

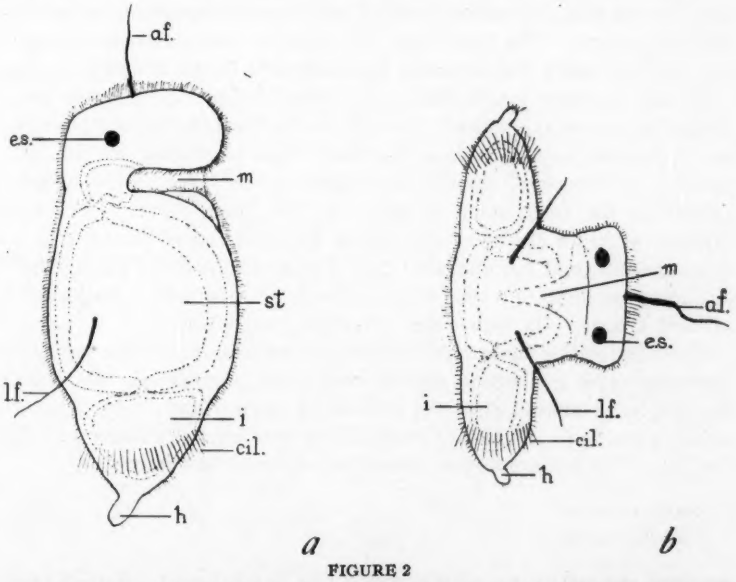


FIGURE 2

Fig. 2(*a*) Normal 6-day trochophore; (*b*) Janus embryo; a.f., apical flagellum; cil., ciliary band; e.s., eye spot; h, holdfast; i, intestine; l.f., lateral flagellum; m, mouth; st., stomach.

difficult, in Chætopterus, to determine whether both are whole embryos. Nevertheless, the number of whole embryos coming from isolated blastomeres in mass cultures makes it fairly certain that each blastomere of the two-cell stage of an equally cleaving egg is capable of producing a whole embryo when isolated.

Double embryos were also obtained by subjecting the eggs to low temperature. When eggs in the "pear-shaped" stage are kept at 10°C. for about an hour, and then returned to room temperature, a few of them cleave equally. Janus embryos result from some of these eggs.

ANODE AND CATHODE SPARKS DIFFERENTIATED BY THE MUCRONATE ELECTRODE

BY CARL BARUS

DEPARTMENT OF PHYSICS, BROWN UNIVERSITY

Communicated January 31, 1928

Anode Points.—The present note is a continuation of the preceding (these PROCEEDINGS, **14,** 1928, p. 50) in which the anode behavior was left outstanding. The method is the same (see Fig. 1), s in atm./10^6, denoting the wind pressure of positive ions when the needle of the electrode protrudes y cm. The spark gap $x = 2$ cm. is just within the power of the machine under the favorable circumstances (point discharge) given.

As the sy anode graph rises more leisurely than the cathode graph, and as the convection current often strikes incidentally before the critical set is reached, experiments of the kind given heretofore are not satisfactory. I, therefore, ignored the critical set in favor of the initial sy graphs for the anode point in figure 3. The three curves (successively raised $s = 20$ for clearness) give much the same run of wind pressure s, when the extrusion y is increased from 0 to about 0.4 cm. The incidental and irregular jump of a convection pressure s to the cusp is indicated by vertical arrows. Its high value is not long sustained.

Notwithstanding the straggling observations (resulting from the purring intermixture of convection current and spark succession), the slope of the graphs is highest when the cathode is earthed and lowest when the anode is earthed. The latter graph starts from zero at a larger y. Thus for $y = 0.12$ cm. successive alternating observations gave

Anode earthed	$s =$	0		0		0	
Cathode earthed	$s =$		30		20		30 etc.,

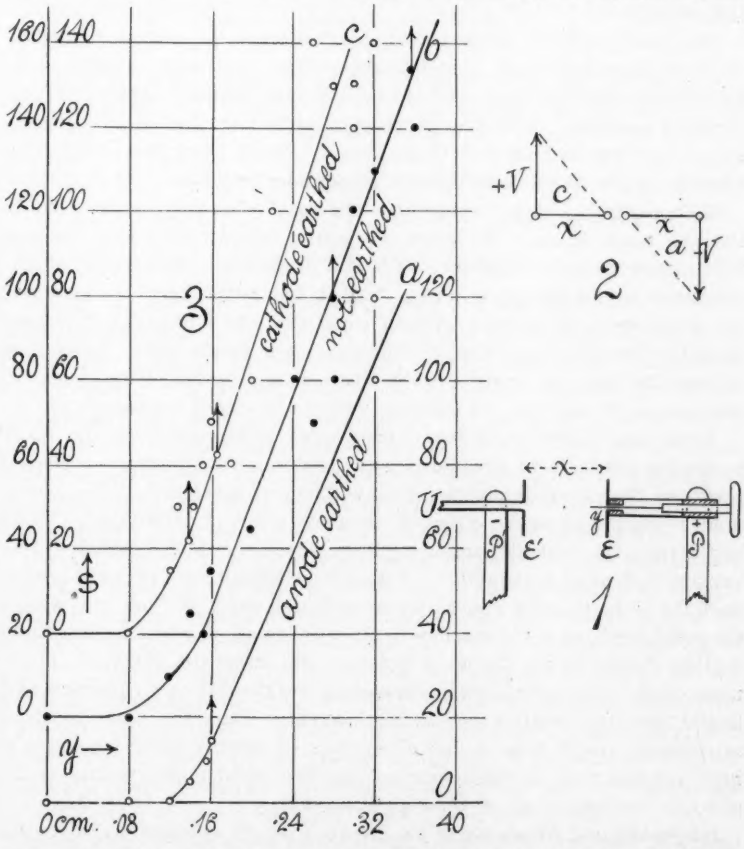
and with the earth connection removed $s = 10$ was found. Similar results were obtained at $y = 0.16$ cm.

Referring to the diagram, figure 2, therefore, the steady emission of positive electrons is greater when the positive potential is high and the negative potential zero than when the positive potential is zero and the negative potential numerically large for the same nominal field. The similarity of these inferences to those in the preceding paper (l. c.) will be noticed, so far as the negative convection current is concerned. Even when the negative needle protrudes, however, sparking is stimulated when its potential is zero. Hence, these sparks appear to be due to positive electrons.

The spark itself at the anode point in these experiments follows an interesting development. From $y = 0$ to $y = 0.10$ cm., the spark gap

(x = 2 cm.) is usually dark with a diffuse glow over parts of the cathode plate.

From y = 0.10 cm. to y = 0.14 cm. straight, brilliantly white linear sparks (darts or arrows) shoot across from the point, each with a sharp snap. The dart of light is fine enough to suggest a dark diffraction line along its middle. But this is an illusion, for under the microscope the line



appeared even in brightness and was measured to be 0.1 mm. broad. The crater of the anode is much finer. The line is, however, nearly uniformly bright longitudinally only from the anode point about 13 mm. towards the cathode; there follows a dark part about 5 mm. long and finally a short bright length again of 2 mm. to the cathode plate. Brightness at the two ends of the dark element is accentuated. From y = 0.14 cm. to 0.30 cm. the individual sparks cease. They pass into a diffuse

cone of purple rays diverging from the anode point and accompanied by a purring noise, showing that many small sparks are in question. An ear trumpet is often a convenience here. This cone decreases in brightness until after $y = 0.30$ is exceeded, when the nearly steady positive convection current begins with maximum s. The spark gap is now quite dark with a faint glow only at the anode point, resembling the cathode in this respect.

The effect produced on the bright line dart when the cathode is earthed (high positive potential) is noteworthy. The dart now extends in an unbroken bright line from anode to cathode and the dark element referred to above is absent. Moreover the inception is earlier, the sparks beginning at $y = 0.06$ cm. instead of 0.10 cm. above. Finally they strike with vigor whereas in the above case there is sometimes hesitancy.

With regard to these occurrences one may tentatively argue that in the line spark or dart, the positive electron is practically shot through from anode point to cathode, with the probability of relaying at the successive atoms along the route; that in the purple cone the electrons are transferred by atomic collision from atom to atom, the ionization spreading laterally by a kind of diffusion; that finally in the convection current the electron transfer is by single atoms moving from anode to cathode, each carrying its electron without essential collision.

Anode and Cathode Sparks Distinguished.—Before the needle point protrudes sparks pass irregularly and promiscuously between preferred points of the electrode plates. This remains unchanged, i.e., the needle point is ignored up to the values of y specified above. After this ($y > 0.06$ cm. about) the cathode emission passes directly into the negative convection current at maximum s. No dartlike sparks were ever obtainable from the cathode in the above apparatus (spark gap $x = 2$ cm. just within the point potential of the machine). It is the anode only which emits these dartlike sparks for the ranges of y values, and under the potential adjustment cited. With increasing y these pass into the diffuse purple cone and finally into the positive convection current. Thus the darts in these experiments result from the forced ejection of positive electrons and it is quite possible that the dark segment near the cathode may terminate the effective ionizing range of these positive rays.

Integration and Resolution of the Electric Dart by a Negative Body.—An interesting observation should be added to the last section. The (positive) dart discharge occurs for a protrusion y of the positive mucronate needle between about $0.10 < y < 0.14$ cm., the limits varying somewhat with the speed of the machine. It is dissipated at the upper threshold of y, into the positive cone, by the presence of the negatively charged rod, within say 10 or 20 cm. of the spark gap, on any side of it, but notably above and on the cathode side. Thus the spark aspect for y greater (incremented)

than the length $y = 0.14$ cm. is evoked by the presence of the negative body. Again, just below the lower threshold, $y = 0.10$ cm., the diffuse spark succession is integrated into the dart by the presence of the negative body. In both cases, therefore, the effect of the latter is equivalent to an increment of y. (Added Feb. 29.)

A SPECTROSCOPIC CRITERION FOR THE BENZENOID STRUCTURE IN SOME TYPES OF TRIPHENYLMETHANE DERIVATIVES[1]

BY R. C. GIBBS AND C. V. SHAPIRO[2]

DEPARTMENTS OF PHYSICS AND CHEMISTRY, CORNELL UNIVERSITY

Communicated January 30, 1927

Except for the general distinction, which is not invariably followed, that color indicates the presence of the quinoid condition in organic compounds, no consistent characteristic or property subject to quantitative measurement has been discovered up to the present time which differentiates benzenoid from quinoid compounds. In the course of an extensive investigation[3] on the absorption spectra of various types of triphenylmethane derivatives, including benzeins, phthaleins, sulfonephthaleins and fluoresceins, a definite regularity has been observed in the case of those compounds known to be benzenoid or lactoid in structure.

In every case, with one possible exception, where the benzenoid structure was to be expected, the absorption spectra of the compounds in neutral alcoholic solution include a pair of bands having an average separation of about 100 mm.$^{-1}$ and lying in the region between frequency numbers (mm.$^{-1}$) 3500 and 3700 as shown in table 1. In general, both bands are relatively weak and are often superimposed on the general absorption. They have approximately the same intensity, although the more refrangible band is frequently of slightly greater intensity. It is to be noted, also, that the average molecular absorption coefficient for these two bands has the same order of magnitude in all the compounds studied.

Application of this criterion to several cases where colored solutions were obtained, gave confirmation to the view that benzenoid or lactoid derivatives were actually present in the solution, even though they might be in equilibrium with colored, quinoid modifications. A striking example is that of the hydration product of fuchsone, *p*-hydroxy-triphenylcarbinol, which has been isolated by Gomberg[4] in both a yellow and a colorless form. When either of these, or fuchsone, is dissolved in ethyl alcohol and the solutions allowed to stand for some time, it is found that all three

approach a common color and a study of their absorption spectra[3d] reveals complete identity in the position of the absorption bands. An equilibrium is established between the colorless carbinol and the yellow quinoid hydrate. Benzaurin and aurin likewise undergo hydration in solution, giving rise to equilibrium mixtures of quinoid hydrates and benzenoid carbinols, and the absorption spectra of their colored hydrated solutions include the characteristic pair of bands listed in table 1. Similar bands are found for alcoholic solutions of the diethyl ethers of phenol- and *o*-cresolsulfonephthalein, which are presumably lactoid but which yield salmon-colored solutions. These examples emphasize the fact that the mere existence of color is not sufficient to warrant the conclusion that the solutes must be in the quinoid state exclusively.

TABLE 1

SUBSTANCE	1ST BAND	2ND BAND	DIFFERENCE	AVERAGE MOLECULAR ABSORPTION COEFFICIENT
Fuchsone[3d]—Hydrated	3515	3610	95	1.68×10^3
Benzaurin[3d]—Hydrated	3520	3613	93	3.16
Aurin[3d]—Hydrated	3514	3628	114	4.93
o-Cresolbenzein[3]	3511	3592	81	4.87
Di-*p*-dihydroxytriphenylmethane[3a]	3571	3695	124	3.45
Diphenylphthalide[3e]	3531	3620	89	1.66
Phenolphthalein[3c]	3523	3613	90	4.61
Isophenolphthalein[3c]	3511	3606	95	5.40
o-Cresolphthalein	3519	3611	92	5.19
Phenol-tetrachlorphthalein	3560	3675	115	5.24
o-Cresoltetrachlorphthalein	3543	3639	96	5.57
Hydroquinolphthalein	3518	3626	108	5.88
Fluoran	3517	3605	88	4.95
Dichlorfluoran	3516	3638	122	4.58
Fluorescein diacetate	3544	3660	116	4.57
Hydroquinolphthalein dibenzoate	3549	3676	127	7.74
Phenolsulphonephthalein diethyl ether	3540	3639	99	4.85
o-Cresolsulphonephthalein diethyl ether	3526	3633	107	7.72

The one apparent exception thus far found to the above rule is dichlorsulphonefluoran, a colorless compound and undoubtedly lactoid in structure. It exhibits a pair of bands at 3238 and 3348, about 300 mm.$^{-1}$ below the position of the bands noted in table 1, with a separation, however, of 115 mm.$^{-1}$, which is comparable with that found for the other benzenoid compounds.

This characteristic pair of bands is not observed in the absorption spectra of neutral, alcoholic solutions of the following compounds, all of which are known to be quinoid:[3] Fuchsone, benzaurin and aurin in freshly prepared solutions, resorcinol-benzein, fluorescein, eosin, phenolsulphonephthalein, *o*-cresolsulphonephthalein and hydroquinolsulphonephthalein.

[1] The investigations upon which this article is based were supported by grants from

the Heckscher Foundation for the Advancement of Research, established by August Heckscher at Cornell University.

[2] Heckscher Research Assistant in Physics and Chemistry, Cornell University, 1923–28.

[3] Orndorff, Gibbs and co-workers, *J. Amer. Chem. Soc.*, (*a*) **47**, 2767 (1925); (*b*) **48**, 1327 (1926); (*c*) **48**, 1994 (1926); (*d*) **49**, 1541, 1545 (1927); (*e*) unpublished data.

[4] Gomberg, J., *J. Amer. Chem. Soc.*, **35**, 1035 (1913).

[5] Orndorff and McNulty, *Ibid.*, **49**, 1588 (1927).

THE DISPERSION BY HYDROGEN-LIKE ATOMS IN UNDULATORY MECHANICS

BY BORIS PODOLSKY

CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated January 30, 1928

1. In view of the recent experimental determination of the dispersion by atomic hydrogen[1] it seems interesting to apply the theory of dispersion developed by Schrödinger[2] to this case. In this paper we restrict ourselves to an approximation in which terms of the order of relativistic correction are neglected. For this purpose it is simpler to obtain our wave equation by the operational method of Schrödinger[3] and Eckart,[4] as extended by Epstein,[5] for in this way we immediately obtain an equation free of relativistic terms.

2. We assume the incident light to be a plane polarized wave of frequency ν propagated along the Y axis, with the electric vector along Z, the nucleus being situated at the origin. The field of the wave and of the nucleus can be represented by a vector potential $\bar{A}$ and a scalar potential ϕ. We take $A_x = A_y = 0$, $A_z = -c^2F \sin \omega(t - y/c)/\omega$, and $\phi = \kappa e/r$, where $\omega = 2\pi\nu$, κe the charge of the nucleus, c the velocity of light and F a constant, all quantities being in electrostatic units.

The corresponding Hamiltonian function is

$$\frac{1}{2\mu}(p_x^2 + p_y^2 + p_z^2) + \frac{eF}{\mu\omega} p_z \sin \omega(t - y/c) + \frac{e^2F^2}{2\mu\omega^2} \sin^2 \omega(t - y/c) - \kappa e^2/r$$

where μ = mass of the electron and $\bar{p} = (p_x, p_y, p_z)$ its momentum. From this we obtain by the method indicated[5] the following equation for ψ

$$\nabla^2\psi - \frac{2ieF}{h\nu} \frac{\partial\psi}{\partial z} \sin \omega(t - y/c) + \frac{8\pi^2\mu e^2}{h^2} \left[\frac{\kappa}{r} - \frac{F^2}{2\mu\omega^2} \sin^2 \omega(t - y/c)\right]\psi - \frac{4\pi i\mu}{h} \frac{\partial\psi}{\partial t} = 0$$

which, to the desired degree of approximation, reduces to

$$\nabla^2\psi + \frac{8\pi^2\mu e^2\kappa}{h^2 r}\psi - \frac{4\pi i\mu}{h}\frac{\partial\psi}{\partial t} = \frac{2ieF}{h\nu}\frac{\partial\psi}{\partial z}\sin\omega t \qquad (1)$$

where on the right we omitted terms in F^2 and a factor $e^{\pm i\omega y/c}$ since $\omega y/c$ is very small for light of visible frequencies and lower.

To solve equation (1) we let

$$\psi = \psi(l, m, n) = A(l, m, n)e^{2\pi i E_l t/h}[\psi_0(l, m, n) + \psi_1(l, m, n)] \qquad (2)$$

where $A(l, m, n)$ is a normalizing factor, and $\psi_0(l, m, n)$ satisfies the equation

$$\nabla^2\psi_0(l, m, n) + \frac{8\pi^2\mu}{h^2}\left(\frac{\kappa e^2}{r} + E_l\right)\psi_0(l, m, n) = 0. \qquad (3)$$

Then, to the desired order of approximation, equation (1) reduces to

$$\nabla^2\psi_1(l, m, n) + \frac{8\pi^2\mu}{h^2}\left(\frac{\kappa e^2}{r} + E_l\right)\psi_1(l, m, n) - \frac{4\pi i\mu}{h}\frac{\partial\psi_1(l, m, n)}{\partial t} = \frac{2ieF}{h\nu}\frac{\partial\psi_0}{\partial z}\sin\omega t. \qquad (4)$$

We now express $2i\sin\omega t$ as the difference of two exponentials; then, letting

$$\psi_1(l, m, n) = e^{i\omega t}u_1(l, m, n) - e^{-i\omega t}u_2(l, m, n) \qquad (5)$$

equating separately the coefficients of the two exponentials, and combining the two resulting equations into one, we obtain

$$\nabla^2 u(l, m, n) + \frac{8\pi^2\mu}{h^2}\left(\frac{\kappa e^2}{r} + E_l \pm h\nu\right)u(l, m, n) = \frac{eF}{h\nu}\frac{\partial\psi_0(l, m, n)}{\partial z} \qquad (6)$$

where, as in the following, the upper sign goes with the subscript 1, the lower with the subscript 2.

3. The next step in our process is to expand $\partial\psi_0/\partial z$ and u each into a series of suitable functions. For this purpose it seems natural to use the set of solutions of the equation of the unperturbed atom, i.e., the set $\psi_0(l, m, n)$, as was done by Schrödinger.[2] Unfortunately, this set is not a *complete* orthogonal set unless a continuous range of complicated functions corresponding to imaginary values of l are included. To avoid this complication we follow a procedure analogous to that used by Epstein for a similar purpose, i.e., we use for our expansion another set of functions, $T(l', m', n')$, defined as follows

$$T(l', m', n') = e^{im'\varphi}P_{n'}^{m'}(\cos\theta)Z(l', n', \alpha) \qquad (7)$$

where $Z(l', n', \alpha)$ satisfies Schrödinger's conditions of finiteness and the differential equation

$$\frac{d^2Z(l', n', \alpha)}{dr^2} + \frac{2}{r}\frac{dZ(l', n', \alpha)}{dr} - \left(\frac{n'(n'+1)}{r^2} + \frac{2\alpha l'}{r} + \alpha^2\right) Z(l', n', \alpha) = 0. \quad (8)$$

We have, therefore,

$$\nabla^2 T(l', m', n') - (\alpha^2 + 2\alpha l'/r)T(l', m', n') = 0. \quad (9)$$

It can be shown[6] that $T(l', m', n', \alpha)$ thus defined, for any constant real value of α, form a complete orthogonal set with respect to a function decreasing rapidly with increasing r, such as $r\partial\psi_0/\partial z$. We shall assume α to be negative. We may, therefore, write

$$r\partial\psi_0/\partial z = \sum a(l', m', n')T(l', m', n') \quad (10)$$

and

$$u = \sum b(l', m', n')T(l', m', n'). \quad (11)$$

It may be objected that since we do not know the properties of the function u we may not write (11), as the set $T(l', m', n')$ may not be complete for this function. However, if to complete the expansion of u we would add to the right member of (11) a sum of terms each of the type cS, where the S's are functions different from the T's, we would find that on account of inhomogeneity of equation (6) all c's must be zeros. Thus the set of T's is sufficient.

Substituting (10), (11) and then (9) into (6) and equating the coefficients of corresponding T's, we obtain

$$b(l', m', n')\left[\alpha^2 + \frac{2\alpha l'}{r} + \frac{8\pi^2\mu}{h^2}\left(\frac{\kappa e^2}{r} + E_l \pm h\nu\right)\right] = eFa(l', m', n')/h\nu r. \quad (12)$$

If we let

$$\alpha^2 = -8\pi^2\mu(E_l \pm h\nu)/h^2 \quad (13)$$

we find that (12) will be satisfied provided

$$b(l', m', n') = \eta eFa(l', m', n')/2h\nu(1 + l'\eta\alpha) \quad (14)$$

where $\eta = h^2/4\pi^2\mu e^2\kappa$.

4. We normalize the solution of (8) in such a way that

$$Z(l', n', \alpha) = \sum_{j=0}^{l'-n'-1} \frac{r^{n'}e^{\alpha r}(2\alpha r)^j}{j!(l'-n'-1-j)!(2n'+1+j)!}; \quad l' > n'. \quad (15)$$

Except for a numerical coefficient and notation these functions are the same as the $\chi(s, \alpha)$ functions used by Epstein,[7] so that we may use the relations obtained by him.

The usual solution of (3) may be put in the form

$$\psi_0(l, m, n) = e^{im\varphi}P_n^m(\cos\theta)\chi(l, n) \tag{16}$$

where

$$\chi(l, n) = \sum_{j=0}^{l-n-1} \frac{r^n e^{-r/\eta l}(-2r/\eta l)^j}{j!(l-n-1-j)!(2n+1+j)!} \tag{17}$$

and

$$E_l = -2\pi^2\mu e^4\kappa^2/h^2l^2.$$

5. We are interested in the case of atoms in the normal state, i.e., the case when $l = 1$, $m = n = 0$. In this case

$$\partial\psi_0(1, 0, 0)/\partial z = -e^{-r/\eta}\cos\theta/\eta$$

so that we can put $m' = 0$, $n' = 1$ and (10) reduces to

$$r\partial\psi_0(1, 0, 0)/\partial z = \cos\theta \sum_{s=2}^{\infty} a(s, \alpha)Z(s, 1, \alpha) \tag{18}$$

where

$$a(s, \alpha) = -\eta^3(1+\alpha\eta)^{s-2}(s+1)!(2\alpha)^4/(1-\alpha\eta)^{s+2}. \tag{19}$$

We can now obtain the expression for $\psi_1(1, 0, 0)$ if we note that now, by (7), (11) and (14)

$$u_1 = \cos\theta \sum_{s=2}^{\infty} b(s, \alpha_1)Z(s, 1, \alpha_1) = \frac{\eta eF}{2h\nu}\cos\theta \sum_{s=2}^{\infty} \frac{a(s, \alpha_1)Z(s, 1, \alpha_1)}{(1+s\eta\alpha_1)} \tag{20}$$

with a similar expression for u_2. In equation (13) we are to take $+$ for α_1 and $-$ for α_2.

If we impose the normalizing condition that $\int \psi(l, m, n)\bar{\psi}(l, m, n)d\tau = 1$ when integrated over the whole space, we obtain for $A(l, m, n)$ of the equation (2) an expression which gives $A^2(1, 0, 0) = 1/\pi\eta^3$.

6. We can now compute the electric moments. We obtain $M_x = M_y = 0$, and

$$\begin{aligned} M_z &= \int z\psi(1, 0, 0)\bar{\psi}(1, 0, 0)d\tau \\ &= -\frac{128e^2\eta^6}{3h\nu} F\cos\omega t\left[\alpha_1^4 \sum_{s=2}^{\infty} \frac{s(s^2-1)(2+s\eta\alpha_1)(1+\alpha_1\eta)^{2s-5}}{(1+s\eta\alpha_1)(1-\alpha_1\eta)^{2s+5}}\right. \\ &\qquad \left. -\alpha_2^4 \sum_{s=2}^{\infty} \frac{s(s^2-1)(2+s\eta\alpha_2)(1+\alpha_2\eta)^{2s-5}}{(1+s\eta\alpha_2)(1-\alpha_2\eta)^{2s+5}}\right]. \end{aligned} \tag{21}$$

Since this quantity M is also the leading term of the matrix $M(l, m, n)$, i.e., $M(1, 0, 0)$, we have for the index of refraction n, the relation $n^2 - 1 = 4\pi MN/F\cos\omega t$, where N is the number of atoms per unit volume.

Thus, we finally have

$$n^2 - 1 = \frac{16Nh^6}{3\pi^5 e^6 \mu^3 \kappa^4 \beta^6}\left[(1-\beta)^2 \sum_{s=2}^{\infty} \frac{s(s^2-1)(2-sq_1)}{(sq_1-1)}\left(\frac{1-q_1}{1+q_1}\right)^{2s} + (1+\beta)^2 \sum_{s=2}^{\infty} \frac{s(s^2-1)(2-sq_2)}{(sq_2-1)}\left(\frac{1-q_2}{1+q_2}\right)^{2s}\right] \quad (22)$$

where $\beta = -h\nu/E_1 = \nu/\nu_1 = h^3\nu/2\pi^2\kappa^2\mu e^4$, $q_1 = \sqrt{(1-\beta)}$, $q_2 = \sqrt{(1+\beta)}$, ν_1 the ionization frequency of the atom.

We may first note that $n^2 - 1$ becomes infinite when $sq_1 - 1 = 0$, or $h\nu = -E_1(1 - 1/s^2)$, i.e., when ν corresponds to one of the absorbtion frequencies of the atom in the normal state. Since q_2 and s are each greater than 1, $sq_2 - 1$ is never zero. Expanding in powers of β and of the wavelength λ we may write

$$n^2 - 1 = \frac{9Nh^6}{32\pi^5 e^6 \mu^3 \kappa^4}(1 + 1.477\beta^2 + 2.39\beta^4 + \ldots)$$

$$= \frac{9Nh^6}{32\pi^5 e^6 \mu^3 \kappa^4}(1 + 1.477c^2/\nu_1^2\lambda^2 + \ldots).$$

When $\nu = 0$, these formulæ give $n^2 - 1 = 9Nh^6/32\pi^5 e^6 \mu^3 \kappa^4$, which is in exact agreement with the result obtained for the dielectric constant by Van Vleck,[8] Epstein[9] and Pauling.[10]

7. For hydrogen we put $\kappa = 1$, and obtain

$$n^2 - 1 = 2.24 \times 10^{-4}(1 + 1.228 \times 10^{-10}/\lambda^2).$$

Substitution of numerical values into (22) gives the following

$h\nu/E_1$	λ IN Å	$(n^2 - 1) \times 10^4$
0.30	3039	2.59
0.25	3647	2.47
0.20	4559	2.38
0.15	6079	2.31
0.10	9118	2.27
0.00	∞	2.24

These results are not in a very good agreement with Langer's determination,[1] but the great experimental difficulties connected with this measurement could account for the disagreement.

I wish to acknowledge my indebtedness to Professor Paul S. Epstein for suggesting this problem and for helpful criticism. My thanks are also due to C. F. Richter and B. Cassen of the California Institute of Technology.

[1] Langer, R. M., *Proc. Nat. Acad. Sci.*, **12**, 639 (1926).
[2] Schrödinger, E., *Ann. Physik*, **81**, 112 (1926), §2.
[3] Schrödinger, E., *Ibid.*, **79**, 745 (1926).
[4] Eckart, C., *Physic. Rev.*, **28**, 711 (1926).

[5] Epstein, P. S., *Proc. Nat. Acad. Sci.*, **12**, 637 (1926), §5.
[6] Courant-Hilbert, *Mathem. Physik*, **V**, §5, p. 239.
[7] Epstein, P. S., *Proc. Nat. Acad. Sci.*, **12**, 630 (1926).
[8] Van Vleck, J. H., *Ibid.*, **12**, 662 (1926).
[9] Epstein, P. S., *Ibid.*, **13**, 432 (1927).
[10] Pauling, L., *Proc. Royal Soc.*, **114**, 198 (1927).

THE PRODUCTION OF ACTIVE NITROGEN

BY JOSEPH KAPLAN[1]

PALMER PHYSICAL LABORATORY, PRINCETON UNIVERSITY

Communicated January 21, 1928

Strutt[2] found that the presence of two per cent of oxygen in nitrogen was sufficient to practically destroy the active nitrogen afterglow. Recently, Constantinides[3] studied the changes produced in the afterglow when foreign gases were introduced into the discharge itself and also when gases were introduced into the afterglow. Among the gas studied was oxygen and the results showed that oxygen caused changes in both the color and in the duration of the active nitrogen afterglow. Both Strutt and Constantinides found that, when the amount of oxygen present was in proportion to the amount present in air, the afterglow produced was a yellow-green afterglow whose spectrum was found to be continuous. Strutt attributed this yellow-green afterglow to the reaction between ozone and nitric oxide. Long exposures on this glow by the author gave only the previously observed continuous spectrum and the OH band at 3064 Å. The absence of band emission indicates that there is no active nitrogen present in this yellow-green glow. As far as the writer is aware no experiments have been reported in which the active nitrogen afterglow has been observed in discharge tubes containing air, and the purpose of this note is to describe briefly some experiments in which active nitrogen has been obtained in air discharge tubes.

The discharge tube was a bulb of about 12 cm. diameter having tungsten electrodes. Air at 5 mm. pressure was pumped through the tube continuously through a capillary open to the atmosphere. No effort was made to exclude either water or mercury vapors from the tube. The discharge was excited by means of an induction coil giving a six-inch spark that corresponds roughly to a voltage of 120,000. The discharge was condensed and a spark gap was connected in series with the tube and the condenser.

The glow was blue in color and changed to the ordinary yellow-green glow with continuous spectrum if the spark gap was excluded. The spectrum of the glow was photographed with a small Hilger quartz spectro-

graph and also with a nitrogen-filled spectrograph, the two instruments embracing the region from 1800 Å to 4500 Å. A rotating sector carrying a commutator on its axle enabled the exposures to be made without danger of light from the discharge tube falling on the slit.

The spectrum of the glow will be discussed fully in a later communication. Besides the α bands of N_2 and the β and γ bands of NO that have previously been observed in the nitrogen afterglow, several new β and γ bands have been identified. Second positive bands of N_2 at 2815, 2953 and 3341 were also definitely identified as well as several members of the fourth positive group of N_2. The number of bands on the plate is so large that the dispersion of the Hilger instrument is too small definitely to separate many bands and some instrument of larger dispersion will have to be used before accurate identification of the bands can be made.

Ruark, Foote, Rudnick and Chenault[4] have reported the presence of second positive group bands in the nitrogen afterglow. The bands which they observed were those at 3577, 3536, 3371 and 3159 Å. For the reason given above the author cannot definitely say that these bands are present on his plates. Bands have, however, been found at every point on the plate where a second or fourth group band should be but because of the overlapping no unique identification of those bands could be made.

As Ruark, et al., pointed out, the presence of the second positive group in the afterglow is of considerable interest because the energy necessary to excite the O-O band of this group is 13.0 volts, which is more than the energy of recombination of nitrogen atoms. The most recent spectroscopic data shows the latter value to be about 11.75 volts. The O-O band of the fourth positive group requires 14.8 volts and the presence of either group in the afterglow raises a difficulty for the hypothesis that active nitrogen is atomic.

That no bands of nitrogen arising from transitions to the normal level of the molecule are present with intensities comparable to the β or γ bands was indicated by the absence of any members of the new Lyman-Birge-Hopfield[5] bands or the predicted A-X[6] bands on the plates taken with the nitrogen-filled spectrograph. Miss Sponer[7] has attempted to photograph the nitrogen afterglow with a vacuum spectrograph and reported negative results with long exposures. The absence of bands due to transitions to the normal state of the molecule is to be kept in mind in any attempts to explain the nature of active nitrogen.

It was thought that the nature of the discharge that excites the active nitrogen afterglow in air would be of some interest and some photographs were taken in order to determine the difference between the active nitrogen producing discharge and the discharge that produces the yellow-green afterglow. The results found are in complete agreement with the results that Strutt[8] obtained when nearly pure nitrogen was used. Strutt

found that the condensed discharge (which produced the active nitrogen) differed from the uncondensed discharge in three striking ways. First, a group of seven bands to which the name of "fourth positive group of N_2" has been given was found to be present in the condensed discharge and absent from the uncondensed discharge. In addition the γ bands of NO, which are present in the uncondensed discharge, were found to be missing from the condensed discharge and finally the second positive group was developed with greater intensity in the condensed than in the uncondensed discharge.

Further experiments showed that the blue active nitrogen afterglow could be obtained with somewhat less intensity, however, when a condensed discharge with spark gap was excited at a voltage of only 10,000. In this case, also, the difference between the condensed and uncondensed discharges was substantially the same as when the higher voltage was used. The spectrum of the condensed discharge in the ultra-violet was, however, not as thoroughly confined to the second and fourth positive groups of nitrogen as when the higher voltage was used. The arc lines of mercury, that were mostly absent under the higher voltage, were strongly developed at the lower voltage.

According to the analysis of the band spectrum of the nitrogen molecule the second and fourth positive groups are emitted when the molecule changes from the C and D levels, respectively, to the B levels. The D level of nitrogen has been found to possess no vibration states since there are no fourth positive bands known that have as their initial state any other than the non-vibrating D level. This seems to show that the nitrogen molecule in the D level is very unstable and Miss Sponer[9] has suggested that the presence of the fourth positive bands in the condensed discharge indicates that active nitrogen is atomic. Both the C and D levels possess more energy than that necessary to dissociate the nitrogen molecule into two normal atoms so that the possibility suggested by Miss Sponer holds as well for the C level. It is possible that more dissociation by collision will occur from the D level than from the C level and this may account for the author's observation of the relatively weak development of the fourth positive group as compared with the increase in the intensity of the second positive group.

[1] NATIONAL RESEARCH FELLOW IN PHYSICS.
[2] Strutt, R. J., *Proc. Roy. Soc.*, **A91**, 303 (1914–15).
[3] Constantinides, *Phys. Rev.*, **30**, 95 (1927).
[4] Ruark, Foote, Rudnick and Chenault, *J. O. S. A. and R. S. I.*, **14**, 1 (1927).
[5] Birge and Hopfield, *Nature*, **116**, 15 (1925).
[6] Birge and Sponer, *Phys. Rev.*, **28**, 259 (1926).
[7] Sponer, *Proc. Nat. Acad. Sci.*, **13**, 100 (1927).
[8] Strutt, R. J., *Proc. Roy. Soc.*, **A85**, 377 (1911).
[9] Sponer, *Zeits. Physik*, **34**, 622 (1925).

ON THE QUANTUM THEORY OF THE RAMSAUER EFFECT

BY J. R. OPPENHEIMER*

NORMAN BRIDGE LABORATORY OF PHYSICS, PASADENA

Communicated February 21, 1928

The Ramsauer effect does not correspond to the decrease in optical scattering by a particle when the wave-length of the light is increased. For one may compute the cross-section for elastic collision of an electron with an atom, without making the approximations used by Born. If one does this, and neglects the resonance of the colliding electron with the atomic electrons one finds that, for atoms with no dipole moment in the normal state, the cross-section approaches a finite limit when the electronic velocity is decreased. The only general result of this calculation is that the distribution of scattered electrons tends, in this limit, to become uniform over all angles.

If, however, one considers the electronic resonance and spin, one obtains results which seem adequate to account for the effects observed by Ramsauer. Thus one may compute two first order cross-sections for the elastic collision of an electron and a hydrogenic atom, corresponding to initial orbital wave functions respectively symmetric and anti-symmetric in the coördinates of the impacting electron and the atomic electron. These turn out to be of the form

$$|f(v, \delta) + g(v, \delta)|^2$$

and

$$|f(v, \delta) - g(v, \delta)|^2.$$

Here v is the electronic velocity, δ the angle of deflection; f and g are positive; f corresponds to transitions in which the atomic electron remains undisturbed, and the free electron changes its direction of motion by the angle δ; g corresponds to transitions in which the free electron takes the place of the atomic electron and this latter electron leaves the atom with a velocity v and a direction of motion δ. For fixed δ and large v, f is much greater than g, while for $v = 0$, $g = 6f$; and f becomes equal to g for velocities of the order of a volt. Thus the symmetric wave function gives a cross-section which is always greater than that computed without resonance: The anti-symmetric one gives a cross-section which, for each angle of deflection, vanishes for some small velocity; and the corresponding total cross-section

$$\int_0^\pi \sin\delta d\delta \, |f(v, \delta) - g(v, \delta)|^2$$

has a sharp minimum at about a volt.

Now, in general, both symmetrical and anti-symmetrical wave functions will occur; and the orbital wave function will involve terms which are symmetrical in the colliding electron and one of the atomic electrons. But if the atom has only paired electrons, the symmetric orbital solution will be excluded by Pauli's principle. Hence in this case (e.g., helium, argon, methane) the cross-section will be given by the anti-symmetric wave function alone, and may pass through a minimum for sufficiently low velocities.

The position and shape of the minimum will depend essentially upon the series of wave functions of the atom, and may be computed only by complete solution of the collision problem; the first order cross-section is inadequate for quantitative prediction. But the occurrence of the minimum depends upon the chemical properties of the atom; this is in good agreement with the available experiments,[1] and particularly with those of Rusch, who reports a sharp minimum in the cross-sections of argon and krypton. Details of the theory will be published elsewhere.

* NATIONAL RESEARCH FELLOW.

[1] *Hdb. d. Phys.*, **23**, 646, 1926.

THE FIFTH DIMENSION IN RELATIVISTIC QUANTUM THEORY

BY NORBERT WIENER AND D. J. STRUIK

DEPARTMENT OF MATHEMATICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Communicated January 24, 1928

O. Klein[1] has shown that the Schrödinger equation

$$\frac{1}{\sqrt{-g}}\sum_{j,k=1}^{4}\frac{\partial}{\partial x^j}\sqrt{-g}\,g^{jk}\frac{\partial\psi}{\partial x^k}+\frac{4\pi ie}{hc}\sum_{j=1}^{4}\varphi^j\frac{\partial\psi}{\partial x^j}-\frac{4\pi^2e^2}{h^2c^2}\left[\sum_{j,k=1}^{4}g^{jk}\varphi_j\varphi_k+\frac{m^2c^4}{e^2}\right]=0 \quad (1)$$

may be cast into the form

$$\frac{1}{\sqrt{-\gamma}}\sum_{j,k=0}^{4}\frac{\partial}{\partial x^j}\sqrt{-\gamma}\,\gamma^{jk}\frac{\partial\Psi}{\partial x^k}=0, \quad (2)$$

where

$$\Psi(x^0, x^1, x^2, x^3, x^4)=e^{\frac{2\pi ix^0}{h}}\psi(x^1, x^2, x^3, x^4),$$

and the matrix γ_{ij} has the components

$$\gamma_{00}=\frac{1}{m^2c^2};$$

$$\gamma_{j0} = \frac{\varphi_j e}{m^2c^3}\ ; \qquad [j = 1, 2, 3, 4]$$

$$\gamma_{jk} = g_{jk} + \frac{\varphi_j\varphi_k e^2}{m^2c^4}\ . \quad [j,k = 1, 2, 3, 4].$$

Here g is the determinant of g_{ij}, γ the determinant of γ_{ij}, g^{ij} and γ^{ij} stand for the appropriate contravariant tensors,

$$\varphi^j = \sum_{i=1}^{4} g^{ij}\varphi_i,$$

c is the velocity of light, e the charge of an electron, m its mass, and x^1, x^2, x^3, x^4 are x, y, z, t. The quantities g_{ij} are the Einstein g's, while the φ_i signify the components of the covariant electromagnetic four-potential. We wish to show that the Klein form of the Schrödinger equation leads to a reasonable interpretation of the function $\psi\bar{\psi}$, which Schrödinger[2] takes to be the electric density and Born[3] to be the probability that an electron lies in a region of space of unit volume (ψ and $\bar{\psi}$ are conjugate complex). We further wish to give a tentative statistical theory of the motions of electrons leading to a function to which we may give this interpretation.

This is perhaps the right place to point out that the earlier interpretations of $\psi\bar{\psi}$ as a space density involve a preferred time direction and are incompatible with a true theory of relativity.[4] The theory to be developed in the present paper is not merely four-invariant, but even five-invariant, at least up to the final stage where $\psi\bar{\psi}$ itself is introduced.

Let $\Phi(x^0, x^1, x^2, x^3, x^4)$ represent the number of electrons at time in a unit volume of space-x^0 ("space-phase"), the latter defined in an invariantive manner. Let us suppose that

$$\Phi(x^0, x^1, x^2, x^3, x^4) = \sum_{n=1}^{N} \psi_n(x^1, x^2, x^3, x^4)e^{i\lambda_n x^0}.$$

Here N is any finite number, and may even be taken to be infinite, if we take proper precautions as to the convergence of the expressions in question. Indeed, our summation may even be over a continuous infinity of frequencies. The "energy" of the component of Φ of x^0-frequency $\lambda_n/2\pi$ will then be given by

$$\psi_n(x^1, x^2, x^3, x^4)\ \psi_n(x^1, x^2, x^3, x^4).$$

We shall regard this component itself as the Ψ of Klein.

Now, it is a familiar notion in optics that two harmonic motions of *nearly* the same frequency may be coherent or incoherent, and that incoherent oscillations have additive energies, not amplitudes. A rigorous

justification of this notion has been given by N. Wiener,[5] on the basis of a more general harmonic analysis than that here introduced. As $\psi\bar{\psi}$ represents a density or a probability, it combines with other quantities of the same sort in accordance with a linear law, even though it is connected quadratically with the solution of a linear partial differential equation. This may be explained by the fact that the different quantities Φ are incoherent in the fifth or phase dimension.

The linearity of $\psi\bar{\psi}$ is thus explained, and we are led to regard Φ itself as the fundamental density or probability, this time in a space-phase of four dimensions. We may confine our attention to real functions, Φ, since our fundamental differential equation (2) is real, and we need not introduce the complex probabilities of Hilbert, v. Neumann and Nordheim,[6] which are related to the complex character of equation (1). Indeed, it is possible in the manner here indicated to give a complete representation of the Hilbert theory. This we reserve for a more detailed paper.

Let it be noted that only one frequency $1/h$ of Φ is involved in the Schrödinger theory. It is more exact to say that only one narrow band of frequencies is involved, as phenomena of incoherence cannot occur in strictly monochromatic oscillatory phenomena. *We thus see a definite reason for introducing a fifth dimension, to give room for this band.* The monochromatic character of phenomena in the fifth dimension may perhaps be a limitation on our observation of the phenomena of nature rather than on these phenomena themselves.

We now turn to the second part of our paper, and try to find a physical model of the wave phenomena represented by Φ. We wish to interpret Φ as a particle density in space-phase, since we desire to retain the discreteness of electrons which is actually observable in the experiments of C. T. R. Wilson[7] and of Millikan.[8] It is not possible to build up an adequate wave theory in which the whole phenomenon is represented by the observable electron density. If this were possible, we should find that interference phenomena would only appear as statistical effects among very large assemblages of electrons. However, there are several interference phenomena in cathode-ray streams in which we can recognize the very scarce individual electrons as they pass by.[9] We may regard the observable density, however, as a rough index of the density of a cloud of unobservable particles of which the observable particles form a more or less uniformly distributed part and a visible indication. We might thus examine the phenomena of sound by exploring a body of slightly ionized gas with an electroscope. The charge of a region of the gas, although an integral multiple of a given quantity, would be a crude measure of its density.[10]

We have taken the theory of sound as a parallel example to our theory. Here a density is the solution of a wave equation. As the theory of sound

may also be founded on the kinetic theory of gases, we are led to look for a statistical five-dimensional theory of particle motion in our case. Now, the kinetic theory of gases is founded on the Maxwell law of distribution of velocities. We wish to pursue the analogy and to base the Schrödinger theory on a law of the distribution of velocities—that is, of directions of particles leaving a point in five-space. In order to be able to specify this law, we shall introduce the new assumption that it involves as little more than the metric of this five-space as possible.

Our fundamental metric γ_{ij} alone determines no such quasi-Maxwellian distribution. It does define a cone of directions at each point, namely, the minimal cone. On the minimal cone itself, it determines no metric and, consequently, no distribution. If we adjoin to the minimal cone C one time-like vector $\mathbf{v}$ at each point, a metric is determined on the minimal cone, for this vector has a conjugate four-space S, and this four-space intersects the minimal cone in an imaginary cone C_1 with a three-dimensional surface. This imaginary cone serves as an absolute in S, and determines an angular metric. We can project hyper-solid angle measurements from S on to C by planes passing through $\mathbf{v}$. Thus $\mathbf{v}$ determines in a well-defined way a metric on C.

Given such a metric, we shall call a distribution of directions on C the normal distribution belonging to $\mathbf{v}$ when of all the particles passing through a region of S of unit four-volume, equally many particles move in equal hypersolid angles. The number of particles passing through a unit hypervolume of S at the vertex of C will be taken as the length of $\mathbf{v}$. Such a distribution corresponds to the superposition of a Maxwellian distribution of velocities and a uniform translation in the case of the theory of sound.

Let us now see what form is taken by the equation of continuity, which says that as many particles leave each point of space-time-phase as enter it. To do this, let us single out a point P, the vertex of the minimal cone C; let us take the direction of $\mathbf{v}$ at P as our time direction and P as our origin of coördinates; and let us have

$$ds^2 = (dx^4)^2 - (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2 \qquad (3)$$

at P. The effects of the curvature of our space-time-phase will be of higher order than those considered in what follows, and will be neglected, so that (3) will hold throughout the region we consider. We shall also neglect certain other effects of higher order.

At a neighboring point, let our distribution vector have the components $(\lambda\epsilon, \mu\epsilon, \nu\epsilon, \omega\epsilon, \rho)$, where $\lambda^2 + \mu^2 + \nu^2 + \omega^2 = 1$, and ϵ is a small quantity. This will correspond to a number of lines per elementary hypersolid angle differing from the number corresponding to $(0, 0, 0, 0, \rho)$ by a factor represented in the limit by

$$1 + \frac{3\epsilon}{\rho} \cos \vartheta,$$

where ϑ is the angle between the projections in space-phase of $(\lambda, \mu, \nu, \omega)$ and the vector determining the elementary hypersolid angle investigated.

Let us write $\mathbf{v}(Q)$ for the value of $\mathbf{v}$ at the point Q with coördinates $(\kappa\alpha, \kappa\beta, \kappa\gamma, \kappa\delta, \kappa)$, where κ is a small quantity and $\alpha^2 + \beta^2 + \gamma^2 + \delta^2 = 1$; and let it have length v and components $(v_0, v_1, v_2, v_3, v_4)$. Then the length of the vector representing the distribution per element of hypersolid angle leaving Q in the direction $(\alpha, \beta, \gamma, \delta, 1)$ will be proportional to

$$\begin{aligned} v + \kappa\alpha \frac{\partial v}{\partial x^0} + \kappa\beta \frac{\partial v}{\partial x^1} + \kappa\gamma \frac{\partial v}{\partial x^2} + \kappa\delta \frac{\partial v}{\partial x^3} + \kappa \frac{\partial v}{\partial x^4} + \\ + 3\kappa \Bigg[\alpha \left\{ \alpha \frac{\partial v_0}{\partial x^0} + \beta \frac{\partial v_0}{\partial x^1} + \gamma \frac{\partial v_0}{\partial x^2} + \delta \frac{\partial v_0}{\partial x^3} + \frac{\partial v_0}{\partial x^4} \right\} + \\ + \beta \left\{ \alpha \frac{\partial v_1}{\partial x^0} + \beta \frac{\partial v_1}{\partial x^1} + \gamma \frac{\partial v_1}{\partial x^2} + \delta \frac{\partial v_1}{\partial x^3} + \frac{\partial v_1}{\partial x^4} \right\} \\ + \gamma \left\{ \alpha \frac{\partial v_2}{\partial x^0} + \beta \frac{\partial v_2}{\partial x^1} + \gamma \frac{\partial v_2}{\partial x^2} + \delta \frac{\partial v_2}{\partial x^3} + \frac{\partial v_2}{\partial x^4} \right\} \\ + \delta \left\{ \alpha \frac{\partial v_3}{\partial x^0} + \beta \frac{\partial v_3}{\partial x^1} + \gamma \frac{\partial v_3}{\partial x^2} + \delta \frac{\partial v_3}{\partial x^3} + \frac{\partial v_3}{\partial x^4} \right\} \Bigg]. \end{aligned}$$

It is then clear that the equation of continuity representing the equality between the number of particles entering and leaving P is

$$\frac{\partial v}{\partial x^4} = \frac{3}{4} \left(\frac{\partial v_0}{\partial x^0} + \frac{\partial v_1}{\partial x^1} + \frac{\partial v_2}{\partial x^2} + \frac{\partial v_3}{\partial x^3} \right).$$

We obtain this equation if we remember that the average values of α, β, γ, δ, $\alpha\beta$, $\alpha\gamma$, $\alpha\delta$, $\beta\gamma$, $\beta\delta$, and $\gamma\delta$ when Q varies over the sphere

$$x^4 = \kappa, \ (x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2 = \kappa^2$$

are all zero, and that the averages of α^2, β^2, γ^2, and δ^2 are each $1/4$.

If we wish to obtain for v the Dalembertian equation

$$\frac{\partial^2 v}{(\partial x^4)^2} = \frac{\partial^2 v}{(\partial x^0)^2} + \frac{\partial^2 v}{(\partial x^1)^2} + \frac{\partial^2 v}{(\partial x^2)^2} + \frac{\partial^2 v}{(\partial x^3)^2}, \qquad (4)$$

we must have

$$\frac{\partial v}{\partial x^\nu} = \frac{3}{4} \frac{\partial v_\nu}{\partial x^4} \qquad [\nu = 0, 1, 2, 3].$$

While it is possible to give various considerations of plausibility which make the form of these equations seem reasonable, it is perhaps more

straightforward to admit that their chief justification is the fact that they permit a further development of the theory. They play the rôle of the equations in the theory of sound which determine the laws of collision.

If the equation (4) is satisfied by v at each point separately, then since it is merely another way of writing (2), this latter equation is satisfied by v throughout space-time-phase. It is to be noted that the precise meaning of the solution differs between different solutions of the equation, in as much as the density it measures corresponds to different choices of time directions. This is, perhaps, not so objectionable as it seems, since the Schrödinger theory also involves the reference of $\psi\bar{\psi}$ to a specific time direction. In any event, in all practical cases, we may suppose that the direction of $\mathbf{v}$ departs but slightly from the time direction corresponding to the $\mathbf{v}$ of a given solution.

With reference to the relation of the present paper to a previous investigation of the authors:[11] Although we there rejected the five-dimensional theory as needless for the purely formal theory of the Schrödinger equation, we reintroduce it here in order to give an account of the phenomenon of incoherency and the atomicity of electricity, which was not at all considered in the previous paper. The normalization theory which we there developed is not employed in the present paper; we regard it, however, as a necessary tool for the further investigation of the problems here considered, more especially if it is desired to reduce the number of the fundamental constants of physics to a minimum. It has a particular bearing on the discussion of the constancy of γ_{00}.[12]

[1] Klein, O., "Quantentheorie und fünfdimensionale Relativitatstheorie," *Zs. Physik*, **37** (1926), pp. 895–906. The views there put forward have, however, partly been retracted by Klein: "Zur fünfdimensionale Darstellung der Relativitatstheorie," *Zs. Physik*, **46** (1927), pp. 188–208.

[2] Schrödinger, E., "Quantisierung als Eigenwertproblem (Dritte Mitteilung)," *Ann. Physik* (4) **80** (1926), pp. 437–490, esp. p. 476; "Quantisierung als Eigenwertproblem (Vierte Mitteilung)," *Ann. Physik* (4) **81** (1926), pp. 109–139, esp. p. 117.

[3] Born, M., "Das Adiabatenprinzip in der Quantenmechanik," *Zs. Physik*, **40** (1926–27), pp. 167–192.

[4] Various methods of avoiding this difficulty have been given by W. Gordon. "Der Compton effekt nach der Schrödingerschen Theorie," *Zs. Physik*, **40** (1926), pp. 117–133; O. Klein, "Elektrodynamik und Wellenmechanik vom Standpunkt des Korrespondenz, prinzips," *Zs. Physik*, **41** (1927), pp. 407–442; P. A. M. Dirac, "The Quantum Theory of the Electron," *Proc. Roy. Soc.*, **A112** (1928), p. 610; H. T. Flint, "Relativity and the Quantum Theory," *Ibid.*, p. 630.

[5] Wiener, N., "The Harmonic Analysis of Irregular Motion," *J. Math. Phys. Mass. Inst. Tech.*, **5** (1926), pp. 99–121 and (second paper), pp. 158–189.

[6] Hilbert, D., Neumann, J. v., und Nordheim, L., "Uber die Grundlagen der Quantenmechanik," *Math. Ann.*, **98** (1927), pp. 1–30.

[7] Wilson, C. T. R., "On the Condensation Nuclei Produced in Gases by the Action of Röntgen Rays, Uranium Rays, Ultraviolet Light, and Other Agents," *Phil. Trans. Royal Soc.*, **192A** (1899), pp. 403–453.

[8] Cf. Millikan, R. A., *The Electron: Its Isolation and Measurement and the Determination of Some of Its Properties*, Chicago, University of Chicago Press, **1917**, 260 pp.

[9] We may cite the recent experiments of Davisson, C., and Germer, L. H., "On the Diffraction of Cathode-Rays by Crystals," *Physic. Rev.*, **30** (1927), pp. 705–741. Equally relevant is the work of F. Kirchner, on the "Compton Effect as Manifested by Individual Electrons," *Ann. Physik*, **83** (1927), pp. 969–76.

[10] Cf. Bohr, N., Kramers, H. A., und Slater, J. C., "Uber die Quantentheorie der Strahlung," *Zs. Physik*, **24** (1924), pp. 69–88. Here a theory of the relation of light quanta to light waves is developed which invites comparison with that developed in the present paper.

[11] Struik, D. J., and Wiener, N., "A Relativistic Theory of Quanta," *J. Math. Phys. Mass. Inst. Tech.*, **7** (1927–28), pp. 1–23.

[12] For a discussion of the constancy of γ_{∞}, see Einstein, A., "Zu Kaluzas Theorie des Zusammenhanges von Gravitation und Elektrizität," *Sitzungsber. Berlin Akad.*, **1927**. Erste Mitteilung, pp. 23–25, 26–30.

ON THE EXTENSION OF THERMODYNAMICS TO GENERAL RELATIVITY

By Richard C. Tolman

Norman Bridge Laboratory of Physics, California Institute of Technology

Communicated January 3, 1928

1. *Introduction.*—The recent interesting article of Lenz[1] on the equilibrium between radiation and matter in Einstein's closed universe makes it desirable to investigate the proper method of extending the ordinary principles of thermodynamics so as to make them hold for considerations in curved space-time where the methods of general relativity must be employed.

The considerations of Lenz were apparently based on principles borrowed with little modification from the ordinary thermodynamics valid in flat space-time. Thus the condition of equilibrium in the universe was assumed to be maximum entropy for variations which did not affect the total energy, and the expressions used for energy and entropy differed from those known to hold in flat space-time merely by the substitution of a quantity representing a volume of curved three-dimensional space in place of an ordinary volume of flat space. Since it is well known, however, that in curved space-time it is, in general, a matter of considerable difficulty to find any suitable expression for energy,[2] and since hitherto, as far as I am aware, no investigations have been made as to a proper general relativity expression for entropy, it is evident that Lenz's treatment does not provide a general nor necessarily a correct method of attacking thermodynamic problems in curved space-time.

It is the purpose of the present article to consider two principles expressed in the form of equations true for all sets of coördinates which may serve as the analogues in general relativity of the ordinary first and second laws of thermodynamics.

In a following article, I then hope to consider the problem of Lenz.

2. *The First Law.*—As the first law for our system of thermodynamics we must evidently take the analogue in general relativity of the law of the conservation of energy. This analogue, which in flat space-time leads to the conservation of energy and also of momentum, is given as is well known by the tensor equation[3]

$$(T_\mu^\nu)_\nu = 0 \tag{1}$$

where T_μ^ν is the total energy tensor for matter and electricity.

In order to make use of equation (1), it will be convenient to change to tensor densities and integrate. To do this, equation (1) can be rewritten in the form[4]

$$\mathfrak{T}_\mu^\nu = (T_\mu^\nu)_\nu \sqrt{-g} = \frac{\partial \mathfrak{T}_\mu^\nu}{\partial x_\mu} - \frac{1}{2}\mathfrak{T}^{\alpha\beta}\frac{\partial g_{\alpha\beta}}{\partial x_\mu} = 0. \tag{2}$$

By integration over a definite four-dimensional region, we then obtain

$$\int\int\int\int \left(\frac{\partial \mathfrak{T}_\mu^\nu}{\partial x_\mu} - \frac{1}{2}\mathfrak{T}^{\alpha\beta}\frac{\partial g_{\alpha\beta}}{\partial x_\mu}\right) dx_1 dx_2 dx_3 dx_4 = 0. \tag{3}$$

This equation is evidently valid for all sets of coördinates and can be used as the starting point for energy (and momentum) investigations.

3. *The Second Law.*—Taking equation (3) as a suitable analogue in general relativity for the ordinary first law of thermodynamics, we must now try to discover a proper expression for the second law of thermodynamics.

Here we are on quite new ground. We have, however, two principles to guide us: In the first place our postulate must be expressible in a form which is true for all systems of coördinates; and in the second place the postulate must be equivalent to the ordinary second law of thermodynamics for the case of flat space-time. These requirements, simple as they seem, are stringent enough to lend considerable probability to any postulate that fulfills them.

To express the postulate that we have chosen as satisfying the conditions we shall first define the *entropy vector* at any given point in space-time by the equation

$$S^\mu = \phi_0 \frac{dx_\mu}{ds} \tag{4}$$

where dx_μ/ds refers to the macroscopic motion of the matter (or energy)

at the point in question and ϕ_0 is the proper density of entropy as measured by an observer moving with that matter.

Taking the divergence of S^μ and multiplying by $\sqrt{-g}$ to prepare for integration we obtain[5]

$$(S^\mu)_\mu \sqrt{-g} = \mathfrak{S}^\mu_\mu = \frac{\partial \mathfrak{S}^\mu}{\partial x_\mu}. \tag{5}$$

If we consider now the integral of (5) over a four-dimensional region which includes the whole of any isolated thermodynamic system,[6] our postulate will be given by the equation

$$\int\int\int\int \frac{\partial \mathfrak{S}^\mu}{\partial x_\mu} dx_1 dx_2 dx_3 dx_4 \geqq 0 \tag{6}$$

under the condition

$$\int\int\int\int \left(\frac{\partial \mathfrak{T}^\nu_\mu}{\partial x_\mu} - \frac{1}{2} \mathfrak{T}^{\alpha\beta} \frac{\partial g_{\alpha\beta}}{\partial x_\mu} \right) dx_1 dx_2 dx_3 dx_4 = 0. \tag{7}$$

4. *Tests of the Postulate.*—We must now examine our postulated equations to see if they meet the necessary requirements.

Looking first at equation (6), we note that $(S^\mu)_\mu$ and $\sqrt{-g}\, dx_1 dx_2 dx_3 dx_4$ are both of them invariants so that the equation itself is a tensor equation of rank zero. Turning now to equation (7), although the integral involved is not a tensor since the integration is taken over a finite region, nevertheless, as shown in the preceding section, the equation is true for all sets of coördinates. Hence, our postulated equations do fulfill the first of the requirements, namely, that of covariance, which we laid down above.

To show that they reduce to the known principles of thermodynamics in flat space-time, we must consider the significance of these equations in the Galilean coördinates which can then be used.

For Galilean coördinates x, y, z, t, we have

$$\sqrt{-g} = 1 \tag{8}$$

and referring to the equation of definition (4), can write for the entropy vector

$$\mathfrak{S}^\mu = S^\mu = \left(\phi_0 \frac{dx}{ds}, \phi_0 \frac{dy}{ds}, \phi_0 \frac{dz}{ds}, \phi_0 \frac{dt}{ds} \right)$$
$$= \left(\phi_0 \frac{dt}{ds} \frac{dx}{dt}, \phi_0 \frac{dt}{ds} \frac{dy}{dt}, \phi_0 \frac{dt}{ds} \frac{dz}{dt}, \phi_0 \frac{dt}{ds} \right). \tag{9}$$

Since, however, on account of its relation to probability, the entropy of an infinitesimal region in flat space-time is known to be an invariant for the Lorentz transformation, entropy density will be so related to the Lorentz-

Fitzgerald factor of contraction ds/dt that we can rewrite equation (9) in the form

$$\mathfrak{S}^{\mu} = (\phi u , \phi v , \phi w , \phi) \tag{10}$$

where ϕ is the density of entropy and u, v and w are the components of macroscopic velocity in the system of coördinates that we are now using.

We then have in these coördinates

$$\frac{\partial \mathfrak{S}^{\mu}}{\partial x_{\mu}} = \frac{\partial}{\partial x}(\phi u) + \frac{\partial}{\partial y}(\phi v) + \frac{\partial}{\partial z}(\phi w) + \frac{\partial \phi}{\partial t} \tag{11}$$

and it is easy to show that the integral of this expression over a three-dimensional isolated volume gives the rate of production of entropy inside the boundary. So that for an isolated system our postulate, equation (6), will reduce in flat space-time to

$$\int_t^{t'} \frac{dS}{dt} dt = (S' - S) \geqq 0 \tag{12}$$

where S and S' are the total entropies of the system at times t and t'.

It is also immediately evident, moreover, from the known value of the energy tensor in flat space-time that equation (7) will reduce for the case $\mu = 4$ to the expression

$$U' - U = 0 \tag{13}$$

where U is the total energy of our isolated system. Equations (12) and (13), however, are the accepted expressions for the thermodynamics of an isolated system in flat space-time. Hence, our postulates also fulfill the second of the two necessary requirements, namely, of reducing to ordinary thermodynamics in flat space-time.

5. *Remarks on the Postulates.*—In conclusion, certain remarks concerning the nature of the postulates will not be out of place.

In order to apply the first law of thermodynamics, as given by equation (3) (7), it will often be more convenient to express the second term of the integrand as a function of the components of the fundamental tensor and its derivatives. This permits us to rewrite the equation in the often more useful form

$$\int\int\int\int \left(\frac{\partial \mathfrak{T}_{\mu}^{\nu}}{\partial x_{\nu}} + \frac{\partial \mathfrak{t}_{\mu}^{\nu}}{\partial x_{\nu}} \right) dx_1 dx_2 dx_3 dx_4 = 0 \tag{14}$$

where $\mathfrak{t}_{\mu}^{\nu}$, the so-called pseudo-tensor of gravitational energy, is a known, though complicated, function of the $g_{\mu\nu}$'s and their differential coefficients. Equation (14) like the equivalent equations (3) (7) is, of course, not a tensor equation, but nevertheless is true in all sets of coördinates.

With regard to the second law of thermodynamics, as given by equation (6), it should be noted that the sign "is greater than" applies to the case of a system which is approaching the condition of thermodynamic equilibrium and implies a difference between "forward" and "backward" directions for the time-like coördinate x_4. The equality sign, on the other hand, may characterize either a system which is in a condition of thermodynamic equilibrium, or one in which changes in the direction of equilibrium are not taking place. To distinguish the important case of thermodynamic equilibrium from one of mere delayed change is usually not difficult, since it is merely necessary to see if zero is the maximum possible value for the integral in equation (6).

[1] Lenz, W., *Physik. Z.*, **27**, 1926 (642–645).

[2] Einstein, A., *Sitz. Ber. Preuss. Akad.*, **1918** (448–459).

[3] See, for example, Eddington, *The Mathematical Theory of Relativity*, Cambridge, **1923**, equations 54.22 and 76.3.

[4] For the reduction see Eddington, l. c., equations 51.31 and 51.51.

[5] For the reduction see Eddington, l. c., equations 51.11 and 51.12.

[6] In the immediately following we shall regard the universe as a whole as an isolated system. I hope to give a further elucidation of this point at a later time.

ON THE VARIATION OF THE ELECTROMOTIVE FORCE IN A PHOTOACTIVE CELL, CONTAINING A FLUORESCENT ELECTROLYTE, WITH THE INTENSITY OF ILLUMINATION

BY WILLIAM RULE*

PHYSICAL LABORATORY, CORNELL UNIVERSITY

Communicated February 4, 1928

Introduction.—When two similar electrodes are immersed in a suitable solution and one of the electrodes is illuminated, the other being kept in the dark, an e.m.f. is produced which varies with the time of illumination. On cutting off the illumination this e.m.f. decreases with time. The effect is well shown using fluorescent solutions such as rhodamine, eosin and fluorescein.

The e.m.f. was first attributed by Goldmann[1] and others to a photoelectric effect, electrons being set free from the photosensitive molecules. Recent work, however, by Perrin[2] and Wood[3] on the photochemical changes produced in fluorescent solutions by illumination, lead one to believe that the e.m.f. owes its origin, in great part, to such photochemical modification of the fluorescent electrolyte as would take place at the illuminated electrode. The e.m.f. can then be attributed to the increase in concentration of the new material or materials formed in the neighbor-

hood of the illuminated electrode, together with the decrease in concentration of the original material. Assuming that the transformation brought about by illumination is irreversible and, therefore, permanent, Russell,[4] working with a solution of rhodamine in absolute alcohol, has recently suggested that the decay of the e.m.f. with time, when the illumination ceases, is due to the diffusion of the changed material away from the neighborhood of the illuminated electrode, and describes experiments in support of this view.

The present paper describes a new type of cell and disposition of apparatus designed to eliminate, as far as possible, the effects of diffusion, and certain experiments which have been carried out on the variation of the effect with intensity of illumination.

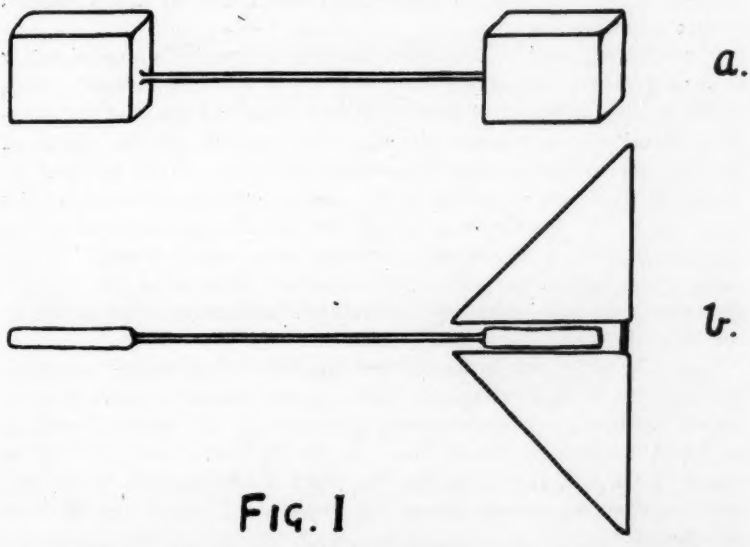


FIG. 1

Apparatus.—The cell is as shown in figures 1*a* and 1*b* and consists of two quartz boxes connected by a fine capillary tube (the cell was made of quartz with a view to possible future needs; the present work was done making use of visible radiation only). The internal dimensions of each box are 9 mm. long, 7 mm. high, and 2 mm. wide; the length of the connecting tube is 30 mm. and its internal diameter about 1 mm. The cell is mounted as shown in figure 1*b*. Light from a Heraeus quartz mercury lamp (110 volts, $2^1/_2$ amp.) is concentrated by means of a thick lens, on the surfaces of two glass right-angled prisms, which latter are disposed about the box to be illuminated in the manner shown. Before reaching the prisms the light passes through five cm. of a solution of

alum in water to absorb the infra-red radiation. Two fine platinum wires, suitably mounted, can be placed centrally, one in each box of the cell, to act as electrodes. The illuminated electrode is connected to one pair of quadrants of an electrometer, the other pair being earthed, together with the unilluminated electrode. The whole of the apparatus is surrounded by an air-tight, earthed metal box, the bottom of which is covered with a portion of the same solution as is used in the experiments to hinder evaporation from the cell. Light passes into this box, and so to the cell, through a glass window, in front of which is a shutter which can be raised or lowered at will. The fine connecting stem of the cell prevents any appreciable diffusion from one box to the other. In order to prevent diffusion to or from the electrode which is to be illuminated, it is necessary to have uniform intensity of illumination throughout, so that a uniform change is produced.

The following consideration shows that this uniformity is approximately realized in a very narrow box when very dilute solutions are used. If I_0 be the incident intensity at each liquid face of the cell box, and the emergent intensity of each beam I_0x^2, then the intensity at the center of the box, due to light passing in one direction, is I_0x. Thus the total intensity at the center of the box is $2I_0x$ and at each liquid face of the box it is $I_0(1 + x^2)$. As the width of the box and the concentration of the solution decrease, x approaches a limiting value which is unity. Thus, using a very narrow box and a dilute solution, the value of the intensity throughout the cell closely approximates to uniformity, approaching in the limit the value $2I_0$.

It is to be noted that in the cell and disposition of apparatus described, the intensity of illumination, and, therefore, the amount of photochemical change produced, is always least at the center of the box and increases gradually towards the boundaries. As the electrode is situated at the center of the box, it is seen that the effect of diffusion of the changed material would be to cause an increase of the e.m.f., when the illumination ceases.

Experimental.—A solution containing 0.05 gram of fluorescein and 10.0 grams of caustic soda per liter of distilled water, was used in the following experiments. A fresh portion of the solution was used for each illumination; the solution itself was kept always in the dark. The platinum electrodes were thoroughly cleaned and when not in use were kept immersed in some of the same fluorescent solution.

Figure 2 gives a curve showing the growth of the e.m.f. with time when one of the boxes is illuminated, and the subsequent behavior of the e.m.f. with time when the incident radiation ceases. During illumination the e.m.f. first of all attains a small positive value which quickly diminishes and, passing through zero, takes on a much greater negative value which

increases with time, finally approaching a maximum value. On cutting off the illumination the e.m.f. has a slight increase, after which it decreases very slowly in the dark.

As any effect of diffusion would tend to increase the e.m.f. when illumination ceases, it would seem that the photochemical change produced is very slowly reversible.

However, as the e.m.f. which attains a value corresponding to 134 scale divisions in one-half hour of illumination, only decays to an extent corresponding to 8 scale divisions in an hour, and thereafter a further 15

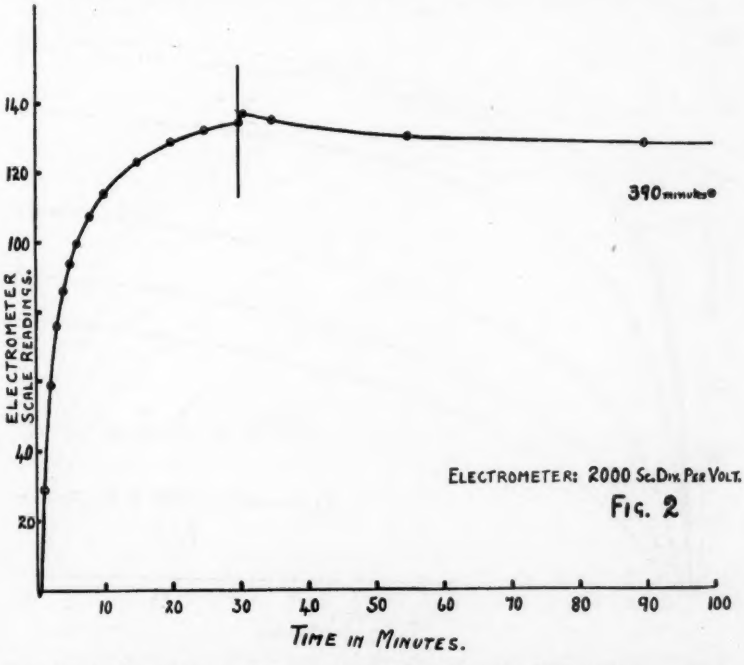


FIG. 2

scale divisions in the subsequent five hours, when illumination ceases, it is evident that the factors which make for decay of the e.m.f. can have but little influence on the form and magnitude of the growth curve, and it is this latter with which we are concerned in this paper.

Experiments were next carried out to determine the symmetry of the apparatus. By covering the face of one of the prisms with black paper the growth curve of the e.m.f. with time was determined when light was passing in one direction through the cell, i.e., when one face of the cell box was illuminated. A similar experiment was performed illuminating the other face of the cell box. The two curves so obtained were identical

in form and very nearly so in magnitude, thus showing the approximate symmetrical disposition of the apparatus. It was noted, however, that the summation of the growth curves so obtained by illuminating each face of the cell box in turn, was much greater than the growth curve obtained when the same illumination is incident on both faces simultaneously.

It would seem, therefore, that there exists a point where increase in the intensity of the incident radiation does not produce any further increase in the magnitude of the effect.

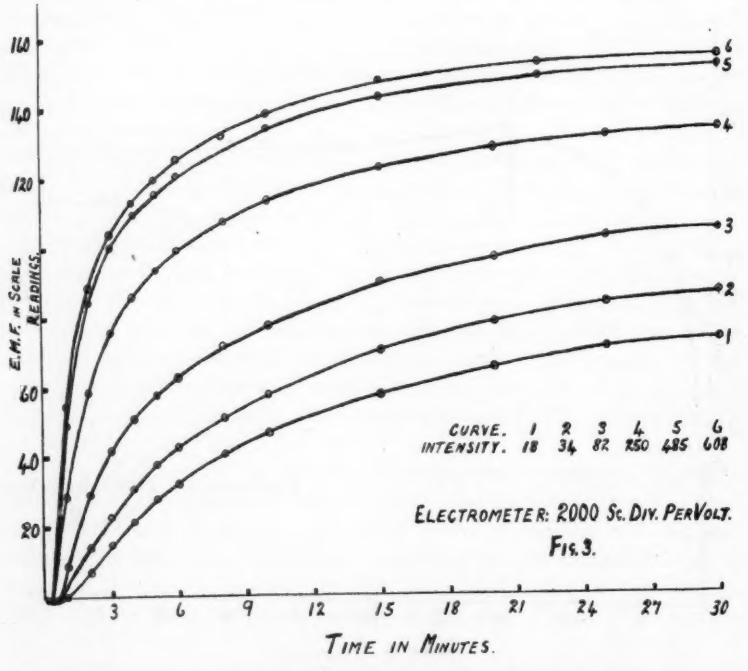


FIG. 3.

Growth curves were then obtained for various intensities of illumination. The intensity of illumination was progressively cut down by placing in the path of the incident beam a series of photographic plates which had been exposed for varying times, and which, therefore, varied in "blackness." Comparative measures of the intensities were made with a sensitive linear thermopile connected to a low resistance galvanometer, the photoactive cell being replaced by the thermopile for such measurements.

Figure 3 gives the family of growth curves obtained in this manner for various intensities of illumination. The fact that a saturation point in the magnitude of the effect is approached with increasing intensity is

evident and is well shown in the family of curves given in figure 4, where the e.m.f. attained after a given time of illumination is plotted against intensity, for several given times of illumination.

It is seen that, in general, for a given time of illumination, the magnitude of the e.m.f. attained increases with increasing intensity in such a fashion as to approach a maximum value for high intensities. Neglecting the small initial positive e.m.f., as the subsequent negative e.m.f. is greatly predominant, there is a marked similarity between the two sets

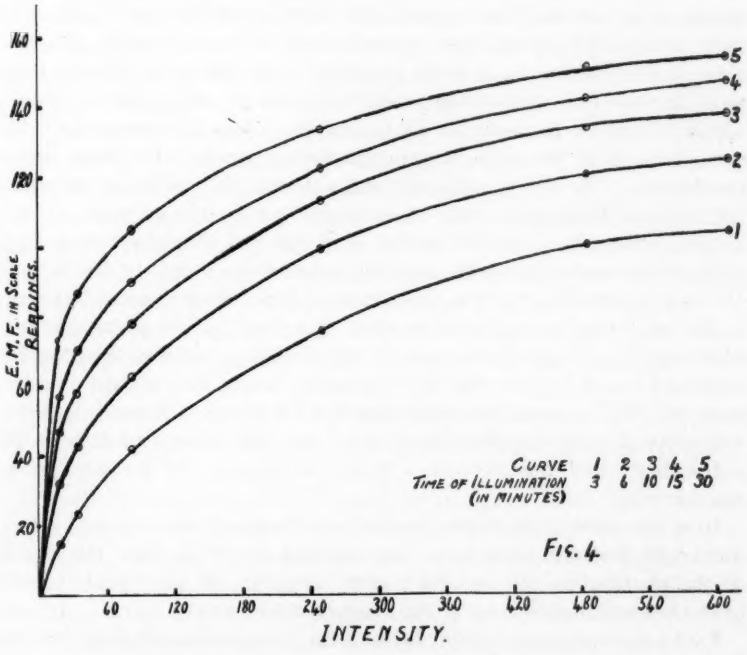


FIG. 4

of curves given, pointing to a general relation which may be expressed as follows:

For any given intensity of illumination, the variation of the negative e.m.f. with time is similar to the variation of this e.m.f. with intensity for any given time of illumination.

Discussion.—Perrin[2] and Wood[3] have shown that when fluorescent solutions, including that used in the experiments here described, are illuminated, a photochemical transformation takes place, the original material gradually disappearing and giving rise to another or possibly other forms. It is thus reasonable to suppose that the e.m.f. produced

in such a photoactive cell as has been described, is in large part due to such photochemical change.

According to the Bunsen-Roscoe law, the rate of change in a photochemical process is proportional to the intensity of the exciting radiation, and the total amount of change is proportional to the product of the intensity and the time of illumination. If the e.m.f. here discussed is due to a photochemical change, it will vary, in some way as yet undetermined, with the amount of change which has taken place. As this latter is proportional to the product of the intensity and the time of illumination, it is seen that, for a given time of illumination, the e.m.f. will vary with intensity in just the same way that the e.m.f. varies with the time of illumination for a given intensity. Also, the curve obtained by plotting the e.m.f. against the product of intensity and time of illumination should be the same for all intensities. The experimental results previously given are only in approximate agreement with these latter conclusions. The divergencies for small values of the product of intensity and time of illumination may be accounted for by the existence of the small positive e.m.f. initially present, and the fact that the long period of the electrometer causes the recorded value of the e.m.f. to lag behind the true value when the rate of change of e.m.f. with time is large, as is the case with the higher intensities. For large values of the product of intensity and time of illumination, the divergencies are in the opposite sense and would indicate that high intensities acting for a certain time are more effective in producing photochemical change than lower intensities acting for a proportionally longer time. In this connection it is to be noted that Wood[3] has recorded a similar violation of the Bunsen-Roscoe law for eosin.

In so far as the experimental results here discussed are in general agreement with photochemical laws, they support the view that the e.m.f. of the photoactive cell described owes its origin, in great part, to the photochemical modification of the fluorescent electrolyte.

I wish to express my indebtedness to the Commonwealth Fund for the Fellowship which has enabled me to study and work in this country, and to Professors Merritt and Murdock for much helpful advice and criticism.

* COMMONWEALTH FUND FELLOW.

[1] Goldmann, *Ann. Physik*, **27**, 449, 1908.

[2] Perrin, *Ann. Physique*, **9–10**, 133, 1918; *Comptes Rendus*, **177**, 612, 1923.

[3] Wood, *Phil. Mag.*, **43**, 757, 1922.

[4] Russell, Ph.D. thesis presented to Faculty of Cornell University, 1927.

A NOTE ON CORONA AT HIGH HUMIDITY

BY A. W. SIMON

TENNESSEE COAL, IRON AND RAILROAD COMPANY, BIRMINGHAM

Communicated February 15, 1928

Recently, during the progress of some experiments on corona, we accidentally made an interesting observation which, while the experiments were not as thorough as desirable, is nevertheless worth recording on account of its important bearing on the problem of thunderstorm electricity.

The apparatus employed was a system of parallel copper wires equally spaced and hung parallel to and equidistant from two parallel steel plates with rounded edges. The voltage furnished by a transformer and synchronous rectifier was applied between the wires and the plates, the wires being negative. The plates were connected to ground through a milliammeter, and a sphere gap was connected in parallel with the system.

We had been making observations on the corona produced by this system for some time previous to December 12, 1927. From December 12th to December 15th heavy rain fell continuously in Birmingham and the observations to be described were made on December 13th and 14th.

A heavy rain was failling at the time, and the apparatus described was set up near an open window. The humidity, therefore, must have been close to 100%.

Under these conditions we found that the corona discharge on the system was singularly altered from that previously observed (on relatively dry days).

Whereas ordinarily, as the voltage was gradually raised, heavy visual corona appeared and persisted up to arc-over, on these two days no corona was visible between the plates, and the current passing from the wires to the plates was enormously reduced. Moreover, on attempting to raise the voltage between the plates, a point was soon reached where a series of intermittent sparks passed across between the electrodes.

These sparks occurred at several points simultaneously, and the location of these points shifted continually. On raising the voltage the number and intensity of these sparks increased. This phenomenon persisted over a considerable voltage range until finally, on raising the voltage still higher, a flaming arc was struck. The voltage at which the spark discharge first appeared was relatively low. The maximum current obtainable previous to the starting of the spark discharge was, in a particular case, 0.3 milliampere as compared to 7.0 milliamperes obtainable just before arc-over on a dry day. Possibly the actual ratio was not as great as this, since we may have lost some current by leakage over the insulating supports, etc. The voltage at which the spark discharge appeared was

35 kv. as compared to 55 kv. for the arc discharge on an ordinary day.

The experiment clearly showed that corona current flow in an atmosphere of high humidity is greatly reduced, and that spark discharges at relatively low voltages are apt to occur. The explanation of this phenomenon very likely is to be found in the reduction of the mobility of negative ions in air of high humidity, although the existence of a thin film of moisture on the electrodes may also play a part.

In this connection it may be pointed out that Elster and Geitel[1] have called attention to the fact that the mobility of gas ions in a fog is greatly reduced; also that Tamm[2] has found that the mobility of gas ions in air at 100% humidity was approximately 25% of that in dry air.

As is well known, with the wires *positive* with respect to the plate instead of *negative*, the corona current flow is greatly reduced and spark discharge sets in at a relatively low voltage even on dry days, so that the action of negative wires on wet days was similar to the action of positive wires on dry days, this is very likely due to the fact that humidity reduces the mobility of the negative ions while it leaves the positive ions unaffected.[3]

[1] *Ann. Phys.*, **2**, 425 (1900).
[2] *Ibid.*, **6**, 259 (1901).
[3] Graetz, *Handbuch der Elek. u. Mag.*, **1923**, Seeliger, III, p. 364.

THE SPARK SPECTRUM OF NEON

BY H. N. RUSSELL, K. T. COMPTON AND J. C. BOYCE

PRINCETON UNIVERSITY

Communicated February 10, 1928

By the use of a vacuum spectrograph in which the spectrum was excited by electron impacts at controlled voltages between a Wehnelt cathode and a wire grid anode, and in which rapid pumping enabled the gas pressure in the spectrograph to be maintained at less than 0.005 of that in the discharge tube, we have discovered an entire new series of lines of considerably shorter wave-length than any hitherto reported for neon. These lines are listed in table 1. That they are the fundamental lines of the neon spark spectrum is suggested by the experimental fact that they are entirely absent at 40-volt excitation, are developed weakly at 60 volts and strongly at 80 volts. This conclusion is confirmed spectroscopically by the appearance in every multiplet of the wave number difference 782 given by the two ionization limits of Ne I. The third, fourth and fifth lines in table 1 are combinations between the two "ground" ^2P states and the three $^4P'$ states discovered by de Bruin (*Zeits. f. Phys.*, **44**, 157,

1927) and serve to connect the new observations with his partial analysis of the near-visible spectrum of Ne II. (Kichlu, in *Proc. Phys. Soc. Lon.*, **39,** 424, 1927, gives several of the quartet terms of de Bruin, but interprets some of them differently.)

TABLE 1

NEW LINES IN SPECTRUM OF NE II

λ	INT.	ν			
462.38	7	216272	$2p$	$^2P_1 - sp^6$	$^2S_1$
460.72	12	217051	$2p$	$^2P_2 - sp^6$	$^2S_1$
456.35	4–	219130	$2p$	$^2P_2 - 3s$	$^4P'_3$
455.27	4	219650	$2p$	$^2P_2 - 3s$	$^4P'_2$
454.68	3	219935	$2p$	$^2P_2 - 3s$	$^4P'_1$
447.83	3	223299	$2p$	$^2P_1 - 3s$	$^2P'_2$
446.60	4	223914	$2p$	$^2P_1 - 3s$	$^2P'_1$
446.26	7	224085	$2p$	$^2P_2 - 3s$	$^2P'_2$
445.05	2	224694	$2p$	$^2P_2 - 3s$	$^2P'_1$
407.17	3	245598	$2p$	$^2P_1 - 3s$	$^2D_2$
405.88	5	246378	$2p$	$^2P_2 - 3s$	$^2D_{32}$
362.55	2	275824	$2p$	$^2P_1 - 3s$	$^2S_1$
361.54	3	276595	$2p$	$^2P_2 - 3s$	$^2S_1$
356.79	2*d*	280277	$2p$	$^2P_2 - 3d$	$^2D_{32}$
353.01	1*d*	283278	$2p$	$^2P_2 - 4s$	$^2P'_{21}$

L. Bloch, E. Bloch and G. Dejardin (*Jour. de Phys.*, **7,** 129, 1926) list 336 lines of Ne II, including some doubtful lines, in the near visible region. With the aid of information given by these new ultra-violet lines and the pioneer work of de Bruin who classified 34 lines in 5 multiplets of the quartet system, we have succeeded in classifying 203 lines in 59 multiplets, thus accounting for all the stronger lines given by Bloch and Dejardin. Table 2 gives the term values of all the states of Ne II thus far identified. The assignment of these levels appears to be assured except those attributed to the 4*f* electron, in which case the reality of the levels is undoubted but their designation is only provisional. X_3 is a fragment of some other multiple level of the sort. All terms have $^3P'$ limits (Ne III) except $3s^2D_{3,2}$ and $3s^2S_2$ associated with ^1D and ^1S terms in Ne III. The ^2D term is apparently an extremely close pair, the corresponding pair in O II differing by only 1 frequency unit.

TABLE 2

TERM VALUES IN NE II

(Terms marked * were discovered by de Bruin)

	QUARTETS			DIFF.	DOUBLETS			DIFF.
					$2p$	$^2P_2$	330429	782
						$^2P_1$	329647	
					sp^6	$^2S_1$	113376	
*	$3s$	$^4P'_3$	111300.0	518.0	$3s$	$^2P'_2$	106343.7	612.2
*		$^4P'_2$	110782.0	299.0		$^2P'_1$	105731.5	
*		$^4P'_1$	110483.0					

TABLE 2 (*Concluded*)

		QUARTETS		DIFF.		DOUBLETS		DIFF.
					$3s$	$^2D_{32}$	84050	
					$3s$	$^2S_1$	53828.5	
	$4s$	$^4P'_3$	48432.6		$4s$	$^2P'_2$	47108.4	
				376.9				571.9
		$^4P'_2$	48055.7			$^2P'_1$	46536.5	
				305.1				
*	$3p$	$^4S'_2$	77477.5		$3p$	$^2S'_1$	77532.6	
*	$3p$	$^4P_3$	84238.2		$3p$	$^2P_2$	76266.0	
				222.8				127.0
*		$^4P_2$	84015.4			$^2P_1$	76139.0	
				182.5				
*		$^4P_1$	83832.9					
*	$3p$	$^4D'_4$	81322.6		$3p$	$^2D'_3$	79419.4	
				337.6				511.0
*		$^4D'_3$	80985.0			$^2D'_2$	78908.4	
				249.7				
*		$^4D'_2$	80735.3					
				144.0				
*		$^4D'_1$	80591.3					
	$3d$	$^4P'_3$	49259.5		$3d$	$^2P'_2$	49098.6	
				−181.8				−859.0
		$^4P'_2$	49441.3			$^2P'_1$	49957.6	
				−221.5				
		$^4P'_1$	49662.8					
*	$3d$	$^4D_4$	51293.5		$3d$	$^2D_3$	50161.9	
				81.0				756.8
*		$^4D_3$	51212.5			$^2D_2$	49405.1	
				106.5				
*		$^4D_2$	51106.0					
				98.2				
*		$^4D_1$	51007.8					
	$3d$	$^4F'_5$	50258.4					
				528.1				
		$^4F'_4$	49730.3					
				97.6				
		$^4F'_3$	49633.7					
				149.4				
		$^4F'_2$	49484.3					
	$4f$	$^4D_4$	27602.0					
				15.0				
		$^4D_3$	27587.0					
				58.9				
		$^4D_2$	27528.1					
				86.7				
		$^4D_1$	27441.4					
	$4f$	$^4F'_5$	27526.7					
				624.4				
		$^4F'_4$	26902.3					
				296.0				
		$^4F'_3$	26606.3					
				−315.4				
		$^4F'_2$	26921.7					
	$4f$	X_3	26830.5					

The term values here given are referred to the $^3P'$ limit in Ne II and are derived from the two series involving the $3s$ and $4s$ electrons.

The mean quantum defects of the various states are given in table 3, where they are compared with those of O II which should, according to

TABLE 3

MEAN QUANTUM DEFECTS OF NE II

	QUARTETS	DOUBLETS	BOTH	VALUES FOR O II
$2p$		0.847	(0.847)	0.724
$3p$	0.672	0.624	0.648	0.548
$3s$	1.015	0.969	0.992	0.867
$4s$	0.991	0.948	0.965	0.857
$3d$	0.045	0.024	0.035	0.030
$4f$	0.010		(0.010)	−0.025

Hund, run closely parallel. Not only are the quantum defects very similar in the two cases, but the term separations are also similar, but usually nearly three times as great for Ne II as for O II.

From the value for $2p^2P_2$, the ionization potential is 40.89 volts. To alter this by 1000 units (0.12 volt) would change the quantum defect for $3d$ by 0.029 and for $4f$ by 0.070 which are well beyond the bounds of probability. The I. P. may, therefore, be taken as *40.9 ± 0.05 volt.* The most accurate estimate of this quantity hitherto available was 38.5 volts, based on extrapolation and assumed parallelism of Moseley curves recently given by Millikan and Bowen (*Phil. Mag.*, **4**, p. 561, 1927).

Note Added to Proof. De Bruin (*Zs. f. Physik*, **46**, 856, 1928) gives additional terms which are fewer than those here tabulated. His arrangement in multiplets differ from ours. We believe our assignment to be more nearly correct—March 7, 1928.

ON HIERARCHICAL CORRELATION SYSTEMS

BY EDWIN B. WILSON

HARVARD SCHOOL OF PUBLIC HEALTH

Communicated January 18, 1928

The correlation system consisting of the $k(k-1)/2$ coefficients r_{ab} existing between $k > 3$ sets $a, b, \ldots$ of n variables $x, y, \ldots$ is said to be hierarchical if the $k(k-1)(k-2)(k-3)/4$ tetrad equations

$$r_{ab}r_{cd} - r_{ac}r_{ad} = 0 \qquad (a, b, c, d, \text{ all different}) \qquad (1)$$

hold.[1] The ordinary way of representing a correlation system graphically is by a scatter diagram of the n points $x_a, x_b, \ldots; y_a, y_b, \ldots; \ldots$ in a space of k dimensions of which one axis is assigned to each of the k sets $a, b, \ldots$. For the present purposes it is more convenient to use a space of n dimensions, one axis for each individual $x, y, \ldots$ and to plot therein k points $a_x, a_y, \ldots; b_x, b_y, \ldots; \ldots$ which represent the values for the individuals for each variable $a, b, \ldots$. These k points may be treated as k vectors $\mathbf{a}, \mathbf{b}, \ldots$ from the origin to the points. If the values are referred to their means as is usual, the k vectors determine a k-space perpendicular to the line $\mathbf{u}$ equally inclined to the n axes (direction cosines all $1/\sqrt{n}$). The length of the vectors may be taken as unity by so choosing the scales of measurement that the standard deviations of all sets are equal to $1/\sqrt{n}$, and the correlation coefficients r_{ab} become the inner products $\mathbf{a.b}$. The correlation system thus comes to be represented by k points on a unit sphere, the cosines of the $k(k-1)/2$ edges are the coefficients, the tetrad relations become

$$\mathbf{a.b\ c.d} - \mathbf{a.c\ b.d} = \mathbf{a} \times \mathbf{b.c} \times \mathbf{d} = 0 \qquad (2)$$

and expresses the fact that in the hierarchical system each edge is perpendicular to all edges not coterminal with it. The conditions may be expressed in terms of the polyhedral angle at the origin determined by the vectors $\mathbf{a}, \mathbf{b}, \ldots$; the plane through any two edges is perpendicular to every plane determined by any two other edges. Such a spherical k-point or polyhedral angle may be called hierarchical.

A plane triangle has an orthocenter, so has a spherical triangle, and a trihedral angle has an orthocentral line such that the plane through it and any edge is perpendicular to the opposite face. In general, a tetrahedron has no orthocenter, the perpendiculars from the vertices to the opposite faces do not meet, similarly a polyhedral angle in four or more dimensions determined by the vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \ldots$ in general has no orthocentric line and the corresponding spherical k-point has no orthocenter. The conditions for an orthocenter are that there is a vector $\mathbf{r}$ such that

$$\mathbf{r} \times \mathbf{a}.\mathbf{b} \times \mathbf{c} = 0 \qquad \text{(for all } \mathbf{a}, \mathbf{b}, \mathbf{c} \text{ which are all different).} \quad (3)$$

We may prove that in this case the figure is hierarchical. To show, for example, that $\mathbf{a} \times \mathbf{b}.\mathbf{c} \times \mathbf{d} = 0$ consider $\mathbf{r} \times \mathbf{a}.\mathbf{c} \times \mathbf{d} = 0$ and $\mathbf{r} \times \mathbf{b}.\mathbf{c} \times \mathbf{d} = 0$ then $\mathbf{r} \times (x\mathbf{r} + y\mathbf{a} + z\mathbf{b}).\mathbf{c} \times \mathbf{d} = 0$, and the whole space of $\mathbf{r}, \mathbf{a}, \mathbf{b}$ is perpendicular to $\mathbf{c} \times \mathbf{d}$ and hence the plane $\mathbf{a} \times \mathbf{b}$ must be. If the plane $\mathbf{r} \times \mathbf{a}$ is perpendicular to $\mathbf{b} \times \mathbf{c}, \mathbf{b} \times \mathbf{d}, \mathbf{c} \times \mathbf{d}, \ldots$ it is perpendicular to the space determined by $\mathbf{b}, \mathbf{c}, \mathbf{d}, \ldots$; but the normal to this space is the polar or reciprocal to $\mathbf{a}$, say $\mathbf{a}'$, and hence $\mathbf{r} \times \mathbf{a}$ passes through $\mathbf{a}'$. Thus the polar polyhedral angle (or spherical k-point) has its edges (or vertices) in the planes (or lines) determined by the orthocentric line (or orthocenter) and the edges (or vertices) of the polyhedral angle (or spherical k-point), and $\mathbf{r} \times \mathbf{a}$ and $\mathbf{r} \times \mathbf{a}'$ are coincident planes. By the reciprocity $\mathbf{b}' \times \mathbf{c}'$ is perpendicular to $\mathbf{a}$ (as $\mathbf{b} \times \mathbf{c}$ is to $\mathbf{a}'$) and hence $\mathbf{r} \times \mathbf{a}'.\mathbf{b}' \times \mathbf{c}' = 0$. The polar polyhedral angle has an orthocentric line which coincides with that of the given polyhedral angle.

Conversely, if the figure is hierarchical there is an orthocenter. Instead of giving a geometric proof one may write down the expression

$$\mathbf{r} = \mathbf{a} - \mathbf{a}' + x\mathbf{a}' \text{ with } \mathbf{a}.\mathbf{a} = \mathbf{a}.\mathbf{a}' = 1, \qquad \mathbf{b}.\mathbf{a}' = \mathbf{c}.\mathbf{a}' = \ldots = 0,$$

and see if x can be so determined that all the relations (3) are satisfied by virtue of (2). We have $\mathbf{r}.\mathbf{a} = x$, $\mathbf{r}.\mathbf{b} = \mathbf{a}.\mathbf{b}$, $\mathbf{r}.\mathbf{c} = \mathbf{a}.\mathbf{c}$, ... For any $\mathbf{b}$ and $\mathbf{c}$

$$\mathbf{r} \times \mathbf{a}.\mathbf{b} \times \mathbf{c} = \mathbf{r}.\mathbf{b}\ \mathbf{a}.\mathbf{c} - \mathbf{r}.\mathbf{c}\ \mathbf{a}.\mathbf{b} = 0$$

and

$$\mathbf{r} \times \mathbf{b}.\mathbf{a} \times \mathbf{c} = \mathbf{r}.\mathbf{a}\ \mathbf{b}.\mathbf{c} - \mathbf{r}.\mathbf{c}\ \mathbf{a}.\mathbf{b} = x\ \mathbf{b}.\mathbf{c} - \mathbf{a}.\mathbf{c}\ \mathbf{a}.\mathbf{b} = 0$$

shows

$$x = \mathbf{a}.\mathbf{c}\ \mathbf{a}.\mathbf{b}/\mathbf{b}.\mathbf{c}.$$

If **c** is replaced by any other vector **d**

$$\mathbf{r} \times \mathbf{b.a} \times \mathbf{d} = \mathbf{r.a\ b.d} - \mathbf{r.d\ a.b} = (\mathbf{a.b}/\mathbf{b.c})(\mathbf{a.c\ b.d} - \mathbf{a.d\ b.c}) = 0.$$

If **b** had been replaced by **d**

$$\mathbf{r} \times \mathbf{d.a} \times \mathbf{c} = \mathbf{r.a\ d.c} - \mathbf{r.c\ a.d} = (\mathbf{a.c}/\mathbf{b.c})(\mathbf{a.b\ d.c} - \mathbf{b.c\ a.d}) = 0.$$

Again $\mathbf{r} \times \mathbf{d.a} \times \mathbf{e} = \mathbf{r.a\ d.e} - \mathbf{r.e\ a.d} = (\mathbf{a.b\ a.c\ d.e} - \mathbf{a.e\ a.d\ b.c})/\mathbf{b.c}$

but $\mathbf{a.c\ d.e} = \mathbf{a.d\ c.e}$ by (2) and hence

$$\mathbf{r} \times \mathbf{d.a} \times \mathbf{e} = (\mathbf{a.d}/\mathbf{b.c})(\mathbf{a.b\ d.e} - \mathbf{a.e\ b.c}) = 0.$$

And $\mathbf{r} \times \mathbf{d.e} \times \mathbf{f} = \mathbf{r.e\ d.f} - \mathbf{r.f\ d.e} = \mathbf{a.e\ d.f} - \mathbf{a.f\ d.e} = 0.$

Thus all the possibilities may be exhausted and the result is proved.

To construct a hierarchical figure we may proceed as follows. First, for the ordinary tetrahedron the argument would be thus: Take any triangle ABC. Mark its orthocenter X and erect a perpendicular XD to the plane. Then the plane DAX is perpendicular to BC, since XD and AX are, and hence the line DA in that plane is perpendicular to BC. As the non-concurrent edges of $ABCD$ are thus orthogonal, the tetrahedron has an orthocenter Y. Erect YE perpendicular to the space $NBCD$ and then EA, EB, EC.ED will be perpendicular to the non-concurrent other edges. For the plane YEA contains the two lines YE and AY perpendicular to BCD and hence every line in it, and in particular EA, must be perpendicular to any lines in BCD. And so on. Owing to the similarity of Euclidean and spherical geometry with respect to perpendicular relations a precisely similar proof holds for the k-point on the $(k-1)$-dimensional spherical surface in k dimensions. How many different hierarchical figures are there? Consider the tetrahedron. The triangle ABC has 9 degrees of freedom, 3 for each vertex, the point D has one degree, making 10 in all. The ∞^{10} tetrahedra reduce to only ∞^4 different ones as rigid motion has 6 degrees of freedom. So in k-space the triangle has 3 k degrees, there are k-2 dimensions left and the perpendicular to the triangle has k-3 degrees of freedom, the fourth vertex has then k-2, the fifth vertex k-3, etc. The total number is $3k + (k-2)-(k-1)/2 = (k^2 + 3k + 2)/2$; but motion has $(k^2 + k)/2$ so that the net number of different k-points is ∞^{k+1}. Similar considerations apply to the spherical k point on the $(k-1)$-dimension surface of the sphere. There are $\infty^{k(k-1)}$ possible k-points of which $\infty^{k(k+1)/2}$ are hierarchical. The loss in degrees of freedom $k(k-1)-k(k+1)/2$ is $k(k-3)/2$ which corresponds[2] to the number of independent conditions among (1). Of the $\infty^{k(k+1)/2}$ hierarchical figures $\infty^{1+k(k-1)/2}$ have a common orthocenter, but the number of rotations about this point are $\infty^{(k-1)(k-2)/2}$ and hence

the number of different hierarchical figures about a given orthocenter is ∞^k—the same number k of degrees of freedom as there are vertices.

If we write

$$\mathbf{r}_0 = \frac{(\mathbf{b.c})\mathbf{a} - (\mathbf{a} \times \mathbf{b.a} \times \mathbf{c})\mathbf{a}'}{\sqrt{\mathbf{a.b}\ \mathbf{a.c}\ \mathbf{b.c}}} = \sqrt{\frac{\mathbf{b.c}}{\mathbf{a.b}\ \mathbf{a.c}}}\,(\mathbf{a} - \mathbf{a}') + \frac{\mathbf{a.b}\ \mathbf{a.c}}{\mathbf{b.c}}\,\mathbf{a}' \qquad (4)$$

$$\mathbf{r}_0.\mathbf{a}\ \mathbf{r}_0.\mathbf{b} = \mathbf{a.b} \qquad \text{(all pairs of different vectors)} \qquad (5)$$

$$\mathbf{r}_0.\mathbf{a} = \sqrt{\frac{\mathbf{a.b}\ \mathbf{a.c}}{\mathbf{b.c}}} \qquad \text{(any } \mathbf{a}, \mathbf{b}, \mathbf{c}, \text{ all different).} \qquad (6)$$

These are nearly the relations Spearman uses in his discussion of general intelligence. To keep things real we have assumed further that all the correlation coefficients **a.b** are positive (or may be made so by appropriate reversals of scales) and that all the first partials $r_{ab.c}$ are also positive so that $\mathbf{r}_0.\mathbf{a}$ is less than unity. (This means that the spherical k-point has edges that subtend acute angles at the center of the sphere and that the angles between edges are also acute.) Under these circumstances $\mathbf{r}_0$ directed from the center of the sphere toward the orthocenter of the spherical k-points must be shorter than the unit vectors **a, b, c,** .... If, then, we take in the space of $n-k-1$ dimensions orthogonal to the k-space and to the equally inclined vector **u,** any vector **v** such that $v^2 + r_0^2 = 1$ we shall have in $\mathbf{g} = \mathbf{v} + \mathbf{r}_0$ a unit vector such that (5) becomes

$$\mathbf{g.a}\ \mathbf{g.b} = \mathbf{a.b}, \qquad \mathbf{g.a} = \sqrt{\frac{\mathbf{a.b}\ \mathbf{a.c}}{\mathbf{b.c}}} \qquad (5', 6')$$

and the components of **g** along the n axes may be taken as the general intelligences of the individuals and from them the specific abilities may be derived.[3] Both the intelligences and the specific abilities are indeterminate, and largely so unless $\mathbf{r}_0$ is very nearly a unit vector. It is clear that the particular linear combination $\mathbf{t} = \alpha\mathbf{a} + \beta\mathbf{b} + \ldots$ of the tests $a, b, \ldots$ which lies along $\mathbf{r}_0$ is that which will correlate highest with **g.**

So long as $\mathbf{r}_0$, which gives the determinate part of **g,** is the same vector in magnitude and in direction, we have the determinate part of the various intelligences unchanged. This will surely be true for the displacement of any hierarchical spherical k-point by rotation about the line $\mathbf{r}_0$, whereby the vectors **a, b,** ... become replaced by certain other vectors $\mathbf{a}_1, \mathbf{b}_1, \ldots$ that are linear combinations of them. Although the new tests give the same intelligences (even the undetermined parts may be taken the same), they will naturally give different specific abilities, namely, those appropriate to the new tests—or new scores derived by weighted combinations of the old scores.[4] But if we are to start upon a system of weighting which leads to new scores we can, while keeping to hierarchical systems,

do pretty much what we want with **g**. We can construct a hierarchical k-point with all the vertices near together which would mean that all the correlations $\mathbf{a}_1.\mathbf{b}_1$, etc., were near to unity, in fact as near as we please. Then **g.a, g.b,** ... would be near to unity and the indeterminate part of g could be reduced to negligible proportions. Moreover, this construction could be carried out in the neighborhood of any point of the sphere that might be chosen as orthocenter, so that if we are to indulge in rearranging our scores by weighting we can make **g** any radius of the sphere, i.e., we can take as **g** any linear combination whatsoever of our original grades. There is ample, almost embarrassing mathematical freedom at our disposal. But mathematics is not psychology, it is logic. It is well for the psychologist to know that logically he may take any weighted scores to represent the general intelligence; he is thus free to concentrate his attention, where he has as a matter of fact generally tried to concentrate it, on the question of agreeing or disagreeing with his colleagues, and others, as to what combination for any particular set of tests should be taken as representing the intelligence, and in how far he desires to leave the intelligence still undetermined, i.e., how long or how short he desires his vector $\mathbf{r}_0$.

With respect to the special abilities which are computed from the equations

$$\sqrt{1 - r_{ag}^2}\,\mathbf{s}_a = \mathbf{a} - r_{ag}\,\mathbf{g} \quad \text{or} \quad \sqrt{1 - (\mathbf{a}.\mathbf{g})^2}\,\mathbf{s}_a = \mathbf{a} - \mathbf{a}.\mathbf{g}\,\mathbf{g}$$

it is clear that for a high degree of determination the values **a.g, b.g,** ... should be very small. From (5′, 6′) this means that the inter-test correlations **a.b, b.c, a.c,** ... should all be small and of about the same order of smallness (so that **b.c** may be large compared with **a.b a.c**), i.e., that the spherical k-point should have nearly quadrantal edges and be approximately regular (equal edges and angles). Under such circumstances **a.g, b.g,**... are all small, **g** is practically eliminated and the scores **a, b,** ... may be taken as the corresponding special abilities. All this, if we may indulge in linear combination, may be accomplished in a great many ways. The problem of the practical psychological engineer is to determine a set of tests for the kinds of different special abilities which he has to measure which shall satisfy the conditions for the elimination of **g**. Herein there is, however, a real statistical difficulty. For the sampling error in **a.b, b.c,** ... is $1/\sqrt{n}$. If there is to be any assurance that **g** has been eliminated it is important that the coefficients should be small yet large compared with $1/\sqrt{n}$. Even if n were 1100 the sampling error would be 0.03, the coefficients should presumably be at least 0.10 which would put **a.g** up around 0.33, $\mathbf{r}_0$ would be small, **g** would consist chiefly of the indeterminate part **v** (here nearly a unit vector) and the special abilities would be indeterminate to about one-third of their standard deviation. More-

over, this indetermination, like all the indetermination above discussed, is measured relative to the variation of the ability among the n persons tested; it does not involve that other kind of indetermination due to the unreliability of the test in assigning the scores to the individuals in the first place, except in so far as this unreliability may bave attenuated the coefficients **a.b, b.c,** ... of correlation between the tests. Although the tetrad equations (1) be independent of the correction for attenuation, so that a hierarchical system remains so, the resolution into **g** and **s**'s does not share the invariance; attenuation tends to diminish $\mathbf{r}_0$, **g.a, g.b,** ... and thus to under-estimate intelligence with over-estimation of special abilities.

If a system be not hierarchical, i.e., if there is overlap between the special abilities, one may by linear combinations reassemble the original scores in such a way as to form a hierarchical system;[5] but the particular manner, of the infinitely many possible, in which this is done will of itself do much to determine not only the independent special abilities but the resulting vector **g.** In particular, the transformation may be to a quadrantal system $\mathbf{a.b} = \mathbf{b.c} = \mathbf{a.c} = \ldots = 0$ which is hierarchical identically and makes all the correlations **a.g, b.g,** ... except one necessarily zero (whether that one also vanishes depends on the way in which the non-quadrantal hierarchical system is allowed conceptually to approach the quadrantal). If it were possible to overcome the statistical difficulty mentioned above, it would seem as though the most hopeful method of attack would be to determine the special abilities by finding uncorrelated tests corresponding to the desired sorts of special ability, tests which, moreover, were uncorrelated with those for intelligence; then, by introducing tests which covered those special abilities severally but lapped over upon the intelligence, we might be left with a certain amount of determination of **g** which could be increased and rendered more determinate by reducing the weight assigned to the special abilities. In view of the statistical difficulty it might be necessary to approach the matter through the careful determination **g** as Spearman appears to suggest. The mathematical, i.e., the logical indetermination of the resolution of a correlation system into hierarchical form with analysis into general and specific terms, seems to imply that whatever significance we come to attribute to the statistical fact that mental tests tend to satisfy the conditions for a hierarchy must lie (1) in the light it sheds on the character of the tests themselves independently of the real existence of general and specific terms or (2) in the scientific postulates that (*a*) there is a general element definable and determinable, at least within certain limits, by an accepted method of finding it out and (*b*) there are different and distinct special abilities s of particular real content definable and determinable independently of the general ability g and of such atomic character that arbitrary linear

combination of them is meaningless. We could apply such methods to discuss the existence and determinability of general and special prosperities using economic indices, or the general and special physiques using anthropometric measures; the fact that our present economic and anthropometric indices did not form a hierarchical system would mean merely that we had not as yet the right indices for the analysis into general and special. The postulates correspond to an almost universal method of speaking, even about properties of matter, of the chemical elements, and indeed suggest the old division into physics and chemistry. Scholasticism is making a strong return, Thomism is much alive; there are certainly human values and validities involved in the resolution into the general and the specials.

[1] For references see *The Abilities of Man* by C. Spearman, 1927, particularly the Appendix. For mathematical treatments see J. C. M. Garnett, *Proc. Roy. Soc. London*, **A96**, 1919, and *Brit. J. Psych.*, **10**, 1920. Garnett assumes throughout that the values of the variables are distributed according to the normal law of error whereas I see no need of such an assumption. I may further refer to my review of Spearman's book in *Science* **67**, 244 (1928).

[2] As proved by Garnett. The difference between my proof and Garnett's lies partly in the disregard of the normal law and partly in the consideration of the space of n-dimensions instead of a space of k-dimensions. I have made no assumption that there is only one additional dimension above the k available, and do not see the appropriateness of such an assumption.

[3] By the equations $\sqrt{1-r_{ag}^2}\,\mathbf{s}_a = \mathbf{a} - r_{ag}\,\mathbf{g}$, $\sqrt{1-r_{bg}^2}\,\mathbf{s}_b = \mathbf{b} - r_{bg}\,\mathbf{g}$, etc. Under these definitions $\mathbf{s}_a.\mathbf{s}_b = \mathbf{a}.\mathbf{b} + r_{ag}r_{bg} - r_{ag}\,\mathbf{b}.\mathbf{g} - r_{bg}\,\mathbf{a}.\mathbf{g} = 0$ so that the $\mathbf{s}$'s are orthogonal (i.e., uncorrelated) and $\mathbf{s}_a.\mathbf{g} = 0$ so that the $\mathbf{s}$'s are all orthogonal to $\mathbf{g}$. Transposing, we have the resolution of the vectors $\mathbf{a}, \mathbf{b}, \ldots$ into linear combinations of $\mathbf{g}$ and $\mathbf{s}_a, \mathbf{s}_b, \ldots$. This would seem to dispose of the criticism in *Nature*, **120**, 1927, p. 181, that the linearity is assumed. Of course, the regressions need not be linear, we are discussing correlation coefficients, not correlation ratios, we have theorems in algebra, nothing about means of arrays in scatter diagrams.

[4] All this is so simple in the n-space of the individuals and seems so difficult to disentangle in the k-space of the tests! Why does Garnett say that the resolution into $\mathbf{g}$ and $\mathbf{s}$'s is unique? Apparently because he is postulating that the nk marks represent only nk special abilities and n general intelligences, i.e., that we can only go up one dimension from k to $k+1$ and that that dimension is specified. It seems to me that there is more room. Indeed, if we add another test we make more room.

[5] To see how this works out in detail consider a correlation table which does not satisfy the tetrad relations. Let three tests be chosen at random but kept constant; call them 1, 2, 3. It is supposed that the coefficients are positive, both zero order and first partials, for these three. Next consider a linear combination of the first three and the fourth as $x_4' = x_4 + \alpha x_1 + \beta x_2 + \gamma x_3$ meaning that for each individual we will replace his score in the fourth test by a score formed by adding to that certain invariable fractions α, β, γ of his scores on the first three tests (we may neglect for the moment the conditions that the dispersion of x_4' shall be unity as that can be brought about later by dividing the new fourth scores as obtained by their standard deviation). To satisfy the tetrad relations we must have $r_{12}r_{34}' = r_{13}r_{24}' = r_{23}r_{14}'$, two independent equations, and we may do this as follows: Multiplying x_4' by x_1, x_2, and x_3 successively and summing over all individuals we have

$$\Sigma x_1 x_4' : \sigma(x_4') = r_{14}' = r_{14} + \alpha + \beta r_{12} + \gamma r_{13}$$
$$\Sigma x_2 x_4' : \sigma(x_4') = r_{24}' = r_{24} + \alpha r_{12} + \beta + \gamma r_{23}$$
$$\Sigma x_3 x_4' : \sigma(x_4') = r_{34}' = r_{34} + \alpha r_{13} + \beta r_{23} + \gamma.$$

Multiplying r_{14}', r_{24}', r_{34}', respectively, by r_{23}, r_{13}, r_{12} and equating we have equations which are linear, and very simple, in α, β, γ; three such equations can be formed but only two are independents and, hence, one of the quantities, α, β, γ may be assumed at random —let it be β. Then if we let

$$\Delta = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix} \text{ and } R_{12} = \begin{vmatrix} 1 & r_{13} \\ r_{23} & r_{12} \end{vmatrix}, \quad R_{35} = \begin{vmatrix} 1 & r_{13} \\ r_{12} & r_{23} \end{vmatrix}$$

the solution is

$$\alpha = \frac{r_{13}r_{24} - r_{23}r_{14} + \beta(r_{13} - r_{12}r_{23})}{r_{23} - r_{12}r_{13}}, \quad \gamma = \frac{r_{13}r_{24} - r_{12}r_{34} + \beta(r_{13} - r_{12}r_{23})}{r_{12} - r_{13}r_{23}}$$

or $$\alpha = \frac{r_{13}r_{24} - r_{23}r_{14} + R_{13}\beta}{R_{23}}, \qquad \gamma = \frac{r_{13}r_{24} - r_{12}r_{34} + \beta R_{13}}{R_{12}}$$

and
$$r_{14}' = \frac{r_{12}r_{13}}{R_{12}.R_{23}}[-r_{14}R_{12} + (1 - r_{13}^2)r_{24} - r_{34}R_{23} + \beta\Delta]$$
$$r_{24}' = \frac{r_{12}r_{23}}{R_{12}.R_{23}}[-r_{14}R_{12} + (1 - r_{13}^2)r_{24} - r_{34}R_{23} + \beta\Delta]$$
$$r_{34}' = \frac{r_{13}.r_{23}}{R_{12}.R_{23}}[-r_{14}R_{12} + (1 - r_{13}^2)r_{24} - r_{34}R_{23} + \beta\Delta].$$

The solution is asymmetric with respect to 1, 2, 3 because the choice of β as arbitrary introduced an element of asymmetry. The tetrads now vanish. Can β be so chosen as to make the new coefficients r_{14}', r_{24}', r_{34}' all positive and the partials involving them also positive? We want

$$\beta\Delta > r_{14}R_{12} - (1 - r_{13}^2)r_{24} + r_{34}R_{23}$$

which can certainly be satisfied by taking β large enough (Δ is positive). Then the partials of the type $r_{14}' - r_{12}r_{24}' = cr_{12}(r_{13} - r_{12}r_{23})$, where c is the common factor

$$c = \frac{1}{R_{12}R_{23}}[-r_{14}R_{12} + (1 - r_{13}^2)r_{24} - r_{34}R_{23} + \beta\Delta]$$

will be positive. As for those of the type $r_{12} - r_{14}'r_{24}' = r_{12}(1 - c^2 r_{12}r_{13}r_{23})$ we must have $c < 1/\sqrt{r_{12}r_{13}r_{23}}$ or

$$\beta\Delta < R_{12}R_{23}/\sqrt{r_{12}r_{13}r_{23}} + r_{14}R_{12} - (1 - r_{13}^2)r_{23} + r_{34}R_{23}.$$

This does not conflict with the previous inequality and can always be satisfied. Thus we get a set of four scores which satisfy our conditions. Then we move on to the fifth, writing x_5' linearly in terms of the four scores x_1, x_2, x_3, x_4' which satisfy the tetrad relations, namely,

$$x_5' = x_5 + \alpha x_1 + \beta x_2 + \gamma x_3 + \delta x_4'/\sigma_4'.$$

We have four constants at our disposal; there are three new independent tetrad relations to be satisfied; we still have a degree of freedom; the algebra and arithmetic become more involved and more tedious but in no wise more difficult in principle. Indeed, if we will trust this process to go on successfully from step to step we have, by

elementary arithmetic and algebra plus this faith in the continued success, ample evidence that the reductions and the freedom indicated in the text are realizable in any numerical case by simple arithmetic; and all that the mathematical proof given in the text accomplishes is to justify our faith in continuing the arithmetical process. It should be noted that β may be so chosen that $c = 0$, in this case all the coefficients $r'_{14}, r'_{24}, r'_{34}$ vanish and the partials are necessarily all positive. In this case all the new tests (subsequent to the first three) $x'_4, x'_5, \ldots$ have the effect of g eliminated and estimate only special abilities.

As an illustration consider the table of anthropological measurements given by Spearman on p. 144 and not satisfying the tetrad relations. Let 1, 2, 3 represent, respectively, height, weight and vital capacity so that $r_{12} = 0.63$, $r_{13} = 0.43$, $r_{23} = 0.57$. The partials are all positive. Now add sitting height as the fourth variable, $r_{14} = 0.805$, $r_{24} = 0.68$, $r_{34} = 0.475$, $\Delta = \begin{vmatrix} 1 & 0.63 & 0.43 \\ 0.63 & 1 & 0.57 \\ 0.43 & 0.57 & 1 \end{vmatrix} = 0.4021$, $R_{12} = 0.63 - 0.43 \times 0.57 = 0.385$, $R_{22} = 0.57 - 0.63 \times 0.43 = 0.359$

$$\beta\Delta > -0.07385 \text{ or } \beta > -0.1837$$

$$\beta\Delta < 0.3495 - 0.07385 \text{ or } \beta < 0.906$$

$$\alpha = -0.4635 + 0.1975\beta, \quad \gamma = -0.0178 + 0.1842\beta.$$

Thus the new variable x'_4 (not yet reduced to unit standard deviation but showing the relative effects of x_1, x_2, x_3) is

$$x'_4 = x_4 + (0.4635 + 0.1975\beta)x_1 + \beta x_2 + (-0.0178 + 0.1842\beta)x_3.$$

If $\beta = -0.18$ $\qquad x'_4 = x_4 - 0.50x_1 - 0.18x_2 + 0.05x_3$

if $\beta = 0.9$ $\qquad x'_4 = x_4 - 0.286x_1 + 0.9x_2 + 0.148x_3$.

Thus to get in that contribution of the sitting height which will make the system consisting of it with height, weight and vital capacity a hierarchical system we must deduct from the sitting height ($\sigma = 1$) a considerable amount of the height ($\sigma = 1$) and add elements from the weight and vital capacity. If we should choose for β its minimum value -0.1837 all the new coefficients $r'_{14}, r'_{24}, r'_{34}$ would vanish and we should have that new variable which showed merely the special kind of physical ability implied by sitting height free of the influence upon it of height, weight or vital capacity. The latter element is unimportant as the figures turn out (correction about 5%). We must deduct between about 50% for the height (because tall people tend to be tall all over or tall whether sitting or standing) and also a correction of about 20% for weight, because for given height heavy people tend to excess sitting height. With this extreme choice of β, nothing is contributed to our knowledge of the general physique by knowing the (new) sitting height. But if we go to the other extreme and take $\beta = 0.9$, then the new sitting height (*ex* effect of height, weight and vital capacity) will make $r_{12} = r'_{14}r'_{24}$, $r_{13} = r'_{14}r'_{34}$, $r_{23} = r'_{24}r'_{34}$ or

$$r'_{14} = \sqrt{\frac{r_{12}r_{13}}{r_{23}}} = r_{1g}, \qquad r'_{24} = \sqrt{\frac{r_{12}r_{23}}{r_{13}}} = r_{2g}, \qquad r'_{34} = \sqrt{\frac{r_{13}r_{23}}{r_{12}}} = r_{3g}$$

and as $r_{ag}r_{bg} = r_{ab}$, $r'_{4g} = 1$. In this case this fourth variable becomes the general physique and consists of the sitting height plus $^9/_{10}$ of the weight plus $^1/_7$ of the vital capacity less $^2/_7$ of the height (all measurements being referred to their standard deviations as units). It would apparently be one of life's handicaps to have any legs at all.

A CORRECTION

Dr. F. Vasilesco has been kind enough to call my attention to an error in my paper "On the Classical Dirichlet Problem for General Domains," which appeared in these PROCEEDINGS, **12**, pp. 397–406 (June, 1926).

Theorem VIII, on page 404, as there stated, is wrong. It becomes correct if the conclusion is changed to read *then the portion of B in any closed region in T' has the capacity 0.*

Theorem VII is correct; moreover, it can be generalized so that it remains the full converse of theorem VIII as corrected. The whole treatment, from theorem VII on p. 403 to the last paragraph on p. 405, can be made satisfactory by changing the property (*b*) of sets *B* to read *the portion of B in any closed region T' has the capacity 0.*

Dr. Vasilesco writes me that he has treated the matter under consideration in a more general form in a paper "Sur les Singularités des Fonctions harmoniques" which will appear in the *Journal des Mathématiques pures et appliqueés* (volume consacré au cinquantenaire scientific de M. E. Picard). It, therefore, seems unnecessary to elaborate further the above correction.

OLIVER D. KELLOGG

Erratum.—Mr. Darcy W. Thompson, of Edinburgh, kindly calls attention to the fact that in the quotation from Newton, page 684, these PROCEEDINGS, **13**, 1927, 3 lines from the end the words "except with accelerative force, they are as the diameters" should read "except with accelerative forces which are as the diameters."

ED.